TECHNISCHE UNIVERSITÄT WIEN

DIPLOMARBEIT

Investigation of the Effect of Complex Raw Materials using Metabolic Flux Balancing - A Case Study on Penicillin Fermentation

ausgeführt zum Zwecke der Erlangung des akademischen Grades
eines Master of Science unter der Leitung von

Prof. Dr. techn. Dipl.-Ing. Christoph HERWIG

betreut durch

Mag. pharm. Dr. Alexandra HOFER

E166 Institut für Verfahrenstechnik, Umwelttechnik und Technische Biowissenschaften

eingereicht an der Technischen Universität Wien
Fakultät für Technische Chemie

von

Stefan F. HAUER, BSc
Matrikelnummer 1125214

Ich erkläre eidesstattlich, dass ich die Arbeit selbständig angefertigt habe. Es wurden keine anderen als die angegebenen Hilfsmittel benutzt. Die aus fremden Quellen direkt oder indirekt übernommenen Formulierungen und Gedanken sind als solche kenntlich gemacht. Diese schriftliche Arbeit wurde noch an keiner Stelle vorgelegt.

Wien, am	eigenhändige Unterschrift	Unterschrift der Betreuerin

With extra time I'd like to think I would've done this better.

Stanley "Fox Stevenson" Byrne

Danksagung

VIELE haben mich in den letzten Jahren auf meinem Weg begleitet und es erst möglich gemacht dass diese Arbeit nach intensiver Arbeit und manchmal noch intensiveren (langjährigen) Pausen zu Stande gekommen ist.

Ein großer Dank geht an meine Eltern, die mich immer unterstützt, aber nie gedrängt haben, und mir die Möglichkeit gegeben haben in Ruhe und meinem eigenen Tempo das Studium abzuschließen.

Ohne den Einsatz von Oliver Spadiut im Biochemielabor im Bachelorstudium hätte es mich vielleicht doch in eine ganz andere Richtung hin verschlagen, und ich hätte nie das Glück gehabt von Daniela Ehgartner in meinem Bachelorpraktikum betreut zu werden. In ihr habe ich eine geduldige Betreuerin gefunden, an die ich mich sowohl am frühen Morgen, als auch spät in der Nacht immer wenden konnte, bei größeren und kleineren Missgeschicken und Unklarheiten.

Professor Christoph Herwig hat mich in seine Forschungsgruppe aufgenommen und mir viele Möglichkeiten gegeben mich zu entfalten, zu wachsen und zu lernen. Danke für den wertvollen Input den du in verschiedene Jour-Fixes gegeben hast, und dass du dich immer um mich bemüht hast.

Ein besonderer Dank geht vor allem an meine "Uni-Eltern" Alexandra Hofer und Paul Kroll, sowie an Julian Kager, Jens Fricke und Ines Stelzer dafür, dass ihr immer ein offenes Ohr für mich hattet und ich jederzeit mit größeren und kleineren Problemen zu euch kommen konnte. Die Arbeit und Diskussionen mit euch haben mir viel Spaß gemacht und vieles für mich erhellt, was sonst im Dunklen verborgen geblieben wäre.

Außerdem möchte ich mich noch bei all den guten Freunden bedanken, die ich im Laufe meines Studiums finden durfte. Ihr habt dafür gesorgt, dass ich mich an der TU Wien immer wie zuhause gefühlt habe und ohne euch wäre diese Zeit nur halb so schön gewesen. Ein Studium ohne Afterworkbier mit Tom, Sophia und Daniel, ohne Beziehungstipps von Donya, ohne Kaffeepausen mit Simona, ohne Programmierdonnerstage mit Lukas und Dorian, ohne Grillen mit Anna, Stefan, Christopher und Babsi, und ohne tiefgründige (und manchmal auch tiefe) Gespräche mit Leo und Vlad - so ein Studium wäre zwar denkbar, aber auch denkbar unspannend.

Zu guter letzt noch eine Nachricht an meine guten Schulfreunde, die nicht mehr daran geglaubt haben, dass ich jemals abschließen werden: Ich habs euch ja gesagt! (Auch wenn ich es manchmal selbst nicht mehr gedacht hätte.)

Zusammenfassung

Die Optimierung des Nährmediums ist oft einer der ersten, aber auch der anspruchsvollsten Schritte in der Entwicklung von Bioprozessen aufgrund ihrer Komplexität. Das führt oft zu Kompromissen in der experimentellen Methodik wie der Verwendung von Schüttelkolben. Diese werden zwar wegen ihrer Einfachheit und der leichten Parallelisierbarkeit verwendet, die Prozessbedingungen sind aber oft weit entfernt von denen in der industriellen Anwendung. In anderen Fällen führt die Komplexität des Prozesses zu einer Reduktion der untersuchten Faktoren basierend auf der Einschätzung von Experten. Dies kann problematisch sein, wenn dieses Wissen lediglich qualitativ ist oder bestimmte Aspekte noch völlig unbekannt sind. Weiters werden aus ökonomischen Gründen oft billige Komplexmedien verwendet, welche aber das Prozessverständnis erschweren und zu zusätzlicher Varianz über Unterschiede in den Chargen führen.

Zwar können Fortschritte in der Analytik viele dieser Probleme lösen, aber bereits jetzt gibt es etablierte mathematische Methoden, welche den Anwender unterstützen können bessere Entscheidungen zu treffen. In dieser Arbeit wird die metabolische Flussanalyse angewandet, um den Einfluss von komplexen Medien auf Fed-Batch-Fermentationen besser zu verstehen. In diesem spezifischen Fall wurde die Produktion des antibiotischen Peptids Penicillin durch *Penicillium chrysogenum* untersucht, in welcher Maisquellwasser als Nährstoffzusatz verwendet wird, um viele hilfreiche Substanzen bereit zu stellen, wie zum Beispiel Aminosäuren. Durch diese Untersuchungen konnte folgendes Wissen etabliert werden:

- Die Berechnung der Raten wird stark dadurch beeinflusst, ob die Freisetzungskinetik der gebundenen Aminosäuren miteinbezogen wird. In einigen Fällen führt dies sogar zur Veränderung des Vorzeichens. Dadurch ändert sich die Interpretation des physiologischen Zustands der Zelle von Grund auf.

- Durch die Methode der metabolischen Flussanalyse konnte nicht nur der Effekt von spezifischen Aminosäuren auf Penicillin- und Biomasseproduktion quantifiziert, sondern auch der Mechanismus dahinter untersucht werden.

Die Vorteile dieses Workflows sind die einfache Anwendung und dessen mechanistische Grundlage. Auf diese Art und Weise kann qualitatives Expertenwissen über den Metabolismus zu einem Model zusammengefügt werden, welches Voraussagen und bessere Entscheidungen in der Prozessoptimierung ermöglicht.

Abstract

MEDIA optimization is often one of the first but also the most challenging steps in bioprocess development due to its sheer complexity. This can lead to compromises in the experimental set up, such as in the use of shake flask experiments. They are used because they are less laborious and easy to run in parallel, however, the conditions inside the shake flasks are often quite different from the ones encountered in the process of interest. In other cases the complexity of the process leads to a restriction of investigated factors based on expert knowledge. This can be problematic if the expert knowledge is only qualitative or if gaps in knowledge exist. Additionally, due to economic reasons, the use of complex media is in many cases preferred, which further complicates process understanding and introduces new uncertainty through lot-to-lot variability.

While advances in analytics and instrumentation might be able to solve many of these problems, currently accessible and well established modelling methods can already guide and help the user in making better decisions based on the available data. Within this thesis the use of metabolic flux analysis is explored as a tool to aid in understanding the role of complex media during a fed-batch process. The specific example chosen was production of the antibiotic peptide penicillin by *Penicillium chrysogenum*, which uses corn steep liquor as media additive to provide many beneficial substances such as amino acids. Based on these investigations, the following could be demonstrated:

- The uptake rates are drastically influenced by incorporation of release kinetics of bound amino acids in corn steep liquor, in some cases even changing the sign of the uptake rates. This fundamentally changes the interpretation of the physiological state of the cell.
- Using metabolic flux analysis, not only the effects of different amino acids on penicillin and biomass formation could be ranked and quantified, with the mechanistic model the mode of action and reasons for different effects could be investigated.

The advantages of the proposed workflow are its easy application, and its well-defined mechanistic basis. In this manner, it can be used for ordering qualitative expert knowledge about metabolism into a model making predictions on possible limitations and improve decision making.

Acronyms

ACV α-aminoadipyl-cysteinyl-valine.

AMC additional media component.

ANN artificial neural network.

BM biomass.

CSL corn steep liquor.

CTR carbon dioxide transfer rate.

DOE design of experiments.

DoR degree of reduction.

HPLC high pressure liquid chromatography.

ICH International Council for Harmonisation.

LP liquid phase.

MCA metabolic control analysis.

MFA metabolic flux analysis.

ODE ordinary differential equations.

OFAT one-factor-at-a-time.

PAT process analytical technologies.

PBP penicillin binding proteins.

PLS partial least squares.

QbD quality by design.

RM raw material.

RNG random numbers generator.

RSM response surface methodology.

SP solid phase.

List of symbols

Below is an alphabetic list of the symbols used in this thesis. If possible, the most common units for each variable are given in brackets, even though in some cases other units might also be possible.

List of greek letters

A	Stoichiometric matrix of substrates S [-]
B	Stoichiometric matrix of products P [-]
Γ	Stoichiometric matrix of biomass constituents X_{Macro} [-]
α	Stoichiometric factor of a substrate S [-]
β	Stoichiometric factor of a product P [-]
δ	Vector of errors on measured fluxes $[mol/mol/h]$
ϵ	Residual vector
θ	Parameter or vector of parameters
θ^*	True parameter or vector of true parameters
$\hat{\theta}$	Estimated parameter or vector of estimated parameters
κ	Condition number
μ	Specific growth rate $[g/g/h]$
σ	True known standard deviation
ϕ	Covariance matrix of the residuals
χ^2	Chi squared distribution
χ^2	Sum of squared residuals

List of latin letters

B^n	Tableau of linear combinations of reversible reactions after the n^{th} iteration
$E(X)$	Diagonal matrix of matrix X
F	Variance-covariance matrix of measured fluxes v_m [-]
F^n	Tableau of linear combinations of irreversible reactions after the n^{th} iteration
G	Stoichiometric matrix of intracellular metabolites
G_c	Stoichiometric matrix of intracellular metabolites containing only the calculated fluxes
G_m	Stoichiometric matrix of intracellular metabolites containing only the measured fluxes
I	Identity matrix
L	Basis of the left null-space
R	Basis of the right null-space
R	Redundancy matrix
R_r	Reduced redundancy matrix
S	Sensitivity
$S^{a,r}$	Absolute-relative sensitivity
S^r	Relative sensitivity
$T^{(n)}$	Tableau to derive the elementary flux modes after the n^{th} iteration

V	Variance-covariance matrix
V	Volume $[L]$
$\dot{V}_{in}$	Volume flow into the system $[L/h]$
$\dot{V}_{out}$	Volume flow out of the system $[L/h]$
X	Biomass $[g]$
$Y_{P/S}$	Yield of product P per substrate S $[g/g]$
c_i	Concentration of substance i $[g/L]$
$c_{i,in}$	Concentration of substance i in $\dot{V}_{in}$ $[g/L]$
g	Sensitivity
k	Michaelis-Menten constant $[g/L]$
q_i	Specific uptake or production rate $[g/g/h]$
r_i	Volumetric reaction rate of substance i $[g/L]$
v	Vector intracellular fluxes of the stoichiometric matrix G $[mol/mol/h]$
v_c	Vector of calculated intracellular fluxes of the stoichiometric matrix G $[mol/mol/h]$
v_m	Vector of measured intracellular fluxes of the stoichiometric matrix G $[mol/mol/h]$

Contents

Part I

Introduction, Goal and Methodical Basis

Chapter 1

Introduction

> There is much pleasure to be gained from useless knowledge.
>
> Bertrand Russell

Penicillium chrysogenum and penicillin

Penicillium chrysogenum sensu lato constitutes several distinct species including *P. chrysogenum sensu stricto*, *P. rubens*, *P. flavigenum*, and many others (Houbraken, Frisvad, and Samson 2011; Scott et al. 2004). The exact classification of the different species within *Penicillium* can be a very difficult task, especially due to misclassifications and reclassifications due to high phenotypic similarities and many historic synonyms: *P. chrysogenum* alone has *P. griseoroseum*, *P. citreoroseum*, *P. brunneorobrum*, and *P. notatum* as synonyms, to note only the most popular. While modern methods make identification easier by using molecular and extrolite analysis, still many species in the genus *Penicillium* await definitive classification (Houbraken, Frisvad, and Samson 2011).

Morphologically, the members of *Penicillium* can be identified by their hyphal growth, their grey-green or grey-blue conidiospores, and their distinct conidiogenous structures (Moss 1987). While *P. chrysogenom* was thought to be completely asexual for over 100 years, the induction of sexual reproduction could recently be demonstrated (Böhm et al. 2013). When grown in submerged culture, *P. chrysogenum* might be described as pellets or in their filamentous form, and may further vary in branching rate (Prosser 1995), the length of filaments (Smith, Lilly, and Fox 1990), pellet size (Nielsen et al. 1995), and vacuolation (Jüsten et al. 1998). These morphological parameters can be affected by agitator shear forces (Smith, Lilly, and Fox 1990; Jüsten et al. 1998; Nielsen et al. 1995), pH and temperature (Miles and Trinci 1983), CO2 concentrations (El-Sabbagh, McNeil, and Harvey 2006; Edwards and Ho 1988), culture age (Makagiansar et al. 1993), and nutrient depletion (McIntyre, Berry, and McNeil 2000).

Just as all fungi, also the members of the family *Penicilium* produce a wide array of hundreds of known secondary metabolites, including ketones, fatty-acid-, shikimate- and amino-acid derivates (Mantle 1987). The most important and prominent are of course the beta-lactams including the different penicillins and cephalosporins. The penicillins affect the penicillin binding proteins (PBP) which play an essential part in cell wall integrity of bacteria. They are able to bind due to their structural similarity to the D-Ala-D-Ala domain of the peptide moiety of peptidoglycan. The high reactivity of the beta-lactam ring system due to high strain leads to the acylation of the PBP and subsequently the thiazolidine ring is able to block further enzymes that would remove the penicilloyl group (Waxman and Strominger 1983; Bycroft and Shute 1987). The beta-lactam antibiotics are therefore classified as so-called ”suicide substrates”.

These antibacterial properties of penicillin were

discovered in 1928 by bacteriologist Alexander Fleming due to a contaminated petri dish (Fleming 1929). While this contamination was initially classified as *P. rubrum* by La Touche, it was later reclassified as *P. chrysogenum* and only recently identified as *P. rubens* by a combination of phylogenetic and extrolite analysis (Houbraken, Frisvad, and Samson 2011). Despite what Fleming would note much later on the origin of this contamination coming through the window from the street, it seems much more plausible that it came from the laboratory of mycologist La Touche downstairs (Hare 1982). This is especially likely since these fungi are thought to be "poorly represented in out-door air in contrast to their indoor frequency" (Scott et al. 2004).

As Fleming seems to have not fully grasped the full potential of his discovery, he envisioned penicillin to be used medically only for superficial infections, but mainly focused on the development of differential media (Hare 1982). It took 10 years and the fortunate discovery of Fleming's publication by Florey and Chain to again set things into motion (Ligon 2004), but the project that followed was of a scale that has been called "second in manpower and cost only to the Manhattan Project" (Bycroft and Shute 1987).

Since then much progress has been made in the industrial production of penicillins. First of all, the natural penicillins, of which over 100 have been identified (Bycroft and Shute 1987), have been supplemented by penicillinase-resistant penicillins, aminopenicillins, and extended spectrum penicillins (Miller 2002). This variety in products is achieved by removing the side chain of a natural penicillin to yield 6-aminopenicillanic acid (6-APA), which can then be modified synthetically (Bycroft and Shute 1987). Through this semi-synthetic approach over a thousand of new penicillins could be discovered and evaluated (Price 1969). Secondly, the industrial strains have been substantially improved. While Flemings original strain would yield the modest amount of 1.2 µg/mL, yields have increased to over 50 000 µg/mL today (Jami et al. 2010). This has been achieved by amplification of the penicillin gene cluster (Newbert et al. 1997), increase of cysteine biosynthesis (Jami et al. 2010), and removal of secondary pathways (Salo et al. 2015). Thirdly, improvements in bioprocess technology have also led to increased biomass production and penicillin yields. Despite the fact that continuous process optimization has been carried out since the 40s, still further increases in yields are reported in the literature (Kiel et al. 2005; Weber et al. 2012). This seems to suggest that even for one of the oldest industrial biotechnological processes there is still a lot of room for further improvements - and therefore also for further research.

However, despite the great importance of antibiotics for health and society, we are amidst a deep crisis, as antibiotic resistances become more common, and pharmaceutical companies have stopped development of novel antibiotics due to high competition and low returns (Fernandes and Martens 2017). Even though bacterial resistances aquired through beta-lactamases, low affinity PBPs, adaption in cell wall proteins, and efflux mechanisms (Finch et al. 2010; Miller 2002; Zervosen et al. 2012) have been a problem described as early as the 40s (Abraham and Chain 1940; Schmidt and Sesler 1943; Schnitzer, Camagni, and Buck 1943), penicillins are still in wide use today. In the decades to come, it will take a tremendous concerted effort on a global scale to stop this emerging threat of antimicrobial resistance. This does not only include the development of new technologies and drugs, but also new policies and a fundamental change in the way resistances are monitored and antibiotics are applied (O'Neill 2016) - it would be a mistake, albeit a common one among scientists, to believe that technology alone would be able to solve our problems.

Complex raw materials in fermentation

DURING development of media composition many factors have to be considered, including process safety, efficiency and cost. Especially for highly competitive products like penicillin, this can lead to the use of complex media components from waste products which will increase growth, and product formation, while at the same

time reducing costs. However, these advantages will often come at the price of increased process uncertainty due to lot-to-lot variability (Zhang et al. 2003) and a generally more complex process which is more difficult to control and optimize. Complex media might introduce many additional questions:

- Which components are present/can be detected in the medium?
- Which media components have a significant effect, positive or detrimental?
- Which media components can even be utilized during the process?
- What lot-to-lot variability is acceptable?

To answer these questions and link the raw material properties to the process outcome, often another problem stands in the way: In many cases appropriate analytical methods do not yet exist and need to be developed first. This does not only increase the workload but also costs time, which is a valuable commodity in industry.

A perfect textbook example of a complex media additive that introduces many problems to process understanding is corn steep liquor (CSL). Not only does it contain a wide range of different substances, including, but not limited to, amino acids, proteins, vitamins (Hofer and Herwig 2017), organic acids, fatty acids, sugars, minerals and heavy metals (Hull et al. 1996), but it is also a multi-phase system consisting of soluble and insoluble components and therefore can prove quite challenging.

CSL is a waste product of corn milling and therefore cheaper than defined media or other complex media as for instance Yeast extract. Due to its varied composition, CSL is used in a wide range of processes for different reasons: In most, it is used as a nitrogen source (P. C. Lee et al. 2000; Kona, Qureshi, and Pai 2001; Gouda, Swellam, and Omar 2001), in others as a carbon (Gudiña et al. 2015), or as a vitamin source (Silveira et al. 2001; Liu et al. 2015).

In industrial and scientific penicillin production by *Penicillium chrysogenum* it is used in the medium of the batch phase, where it increases biomass and penicillin yield (Moyer and Coghill 1946). The importance of the discovery of this effect has been quite enthusiastically noted by Coghill, who said that "one of the least understood miracles connected with [penicillin] is that Florey and Heatley were directed to our laboratory in Peoria—the only laboratory where the corn-steep liquor magic would have been discovered." (Bickel 1972) Several different hypotheses for its effect have been postulated, among them reduced energy consumption due to directly providing metabolites, and an improved growth due to mixed nitrogen sources (Calam and Ismail 1980).

Metabolic flux analysis

BASED on an idea by Verhoff and Spradlin (1976) metabolic flux analysis (MFA) originated during the 70s (Çalık and Özdamar 2002) from a publication by Aiba and Matsuoka (1979). The concept was later extended by researchers such as Papoutsakis and Meyer (1985) and Papoutsakis (1984) and is based on the principles of linear algebra. A stoichiometric matrix is created where each column denotes a single reaction within this biochemical reaction network, and each row is therefore the material balance for a single metabolite in this network (Antoniewicz 2015). If a pseudo steady state is assumed, i.e., internal metabolite concentrations do not change, and metabolite dilution due to growth can be ignored, then this leads to the conclusion that material balances must close and be zero (Stephanopoulos, Aristidou, and Nielsen 1998). Depending on the dimensions of the reaction network, if enough fluxes are measured, the system will be determined or (preferably) overdetermined and the unknown fluxes can easily be calculated. Alternatively, flux balance analysis returns a solution space for underdetermined systems by assuming boundaries for fluxes and maximising certain fluxes, e.g., biomass or ATP (Feist and Palsson 2010).

While initially used to compare the likelihood of the activity of different metabolic pathways (Aiba and Matsuoka 1979), the concept has been applied to a wide range of other problems. These include

but are not limited to strain characterization (Harris et al. 2000), deduction of unknown stoichiometric factors (Henriksen et al. 1996), identification of alternate pathways, soft sensor (Ohadi, Legge, and Budman 2015), calculation of maximum possible yields (Van Gulik et al. 2000), and media optimization (Xie and Wang 1994b; Xie and Wang 1994a).

The use of MFA has several advantages: It is a sound and tried methodology based on well-established mechanistic principles and therefore does not treat the cell as a black box. The primary metabolism has, in rough strokes, been elucidated decades ago, and information on pathways present in different organisms are freely available in specialised databases (Ogata et al. 1999; Caspi et al. 2006). Successfully established MFA models can be found in the literature for simple bacteria as *E. coli* (Long and Antoniewicz 2014), just as for plants (Schwender 2008), fungi (Nissen et al. 1997) and cell culture (Galleguillos et al. 2017). In contrast to other techniques such as metabolic control analysis and biochemical systems theory, no kinetic information on the network is needed (Stephanopoulos and Vallino 1991). The only inputs needed, are the extracellular uptake and production rates, which are calculated for fermentations as a standard procedure. Therefore, the additional experimental burden is quite limited.

However, there are also some limitations and caveats one should keep in mind while applying MFA. As this method assumes a steady state, in the strictest sense it is meant to be applied to fermentations inside a chemostat where this presumption holds true. When applied to inherently dynamic processes, such as batch and fed-batch fermentations, it is argued that metabolic pseudo steady state is achieved within seconds or minutes, and therefore MFA can be applied by calculating rates for distinct phases within the process (Antoniewicz 2015). Others achieve dynamic metabolic flux analysis by fitting extracellular data and using the first order derivative as an input (Niklas et al. 2011), while some model the macro-reactions through kinetic expressions following simple Monod kinetics (Provost and Bastin 2004; Nolan and K. Lee 2011). Another limitation of metabolic flux balancing is that it is not suited for elucidating parallel pathways, cyclic pathways, and reversible reactions for which the more laborious and complicated isotope-based flux analysis is needed (Wiechert 2001). Additionally, it has been suggested that ^{13}C-labelling is also superior to pure metabolite balancing because it "allows the elimination of doubtful cofactor balances", however, this method is not without its own problems and pitfalls, such as assumptions of reversibility of reactions and metabolic isotope effects (Winden, Verheijen, and Heijnen 2001).

Corn steep liquor in penicillin fermentation

THE understanding of complex media components in a microbial process can seem a daunting task. Several different statistical and computational methods have been used to aid in this task. While one-factor-at-a-time (OFAT) methods have fallen out of favour due to being very laborious and not being able to guarantee to find the true optimum, response surface methodology (RSM) used on design of experiments (DOE) have become well established (Beg, Sahai, and Gupta 2003; Desai et al. 2008). Additionally, more modern approaches, like artificial neural network (ANN), have successfully been applied to optimisation of fermentation media (Desai et al. 2008). However, all these methods share a common flaw: They use statistical and computational methods, but do not utilise mechanistic process knowledge.

It has recently been noted by Singh et al. (2017) that MFA might be a valuable method for better understanding how microorganisms utilize medium. The reason for this is twofold: i) By giving mechanistic insights into the metabolism of our organism, we may be able to notice shifts in metabolism due to limitations and devise new fermentation strategies based on these observations. ii) By being able to realise that a specific pathway or substrate might be underutilized we can change this limitation by removing this metabolic bottleneck by genetic engineering. While these suggestions have already

Table 1.1: Overview of published metabolic flux models for *P. chrysogenum*. Other (more recent) published models were excluded if they utilised ^{13}C-labeling.

Publication	Nodes	Fluxes	Structured	Rank	Condition	v_m	CSL
Jørgensen, Møllgaard, et al. (1992)[1]	35	37	No	-	-	9	Yes
Jørgensen, Nielsen, et al. (1995)[2]	50	61(+21)	Yes	50	307	33	Yes
Henriksen et al. (1996)	64(+13)	72	Yes	64	484	8	No
Van Gulik et al. (2000)[3]	<216	193(+2)	Yes	-	-	24	No

[1] As the model is not described in the publication, rank and condition could not be evaluated.
[2] In the publication the amount of intracellular components is mentioned as 49. Even after several checks, it could not be verified if this is a mistake in the publication, or a mistake in the reproduction of the model.
[3] The description of the model is not sufficient to exactly recreate the model used in this study. While all reactions are given, many metabolites in the reactions seem to be there only for completeness, e.g., "H_2O:cyt", "H:cyt", while most definitely not appearing in the model.

been made long before (Varma and Palsson 1994) and some researchers have tried to apply this idea (Xie and Wang 1994b; Xie and Wang 1994a; Xing et al. 2011), it seems that little progress has been made in the last 30 years towards a mechanistic, quantitative optimization procedure that properly utilizes the capabilities of MFA.

There exist several publications that have tried to modify amino acid concentrations of media based on MFA and have been able to successfully improve biomass and product yields (Xie and Wang 1994b; Xie and Wang 1994a; Xing et al. 2011). However, these approaches, besides being far and wide between, have still several shortcomings:

- They use the calculated internal fluxes to interpret the measured uptake rates, but do not quantify the relative effects of the amino acids and also do not try to predict the change in yields based on modified uptake rates. Especially for more complex metabolic networks, this limits the applicability of MFA in a wider scale and its widespread use in research and industry.
- While results of modified media were later tested on fed-batch cultivations, the MFA model itself was tested on chemostat-cultures. This does not take into account that conditions and demands of the cell might change over time during the process, and different process stages might be limited by different amino acids.

Another approach by Gheshlaghi et al. (2007) investigates the effect of amino acids on batch fermentations of *Aspergillus niger*. They use an underdetermined metabolic model with biomass production as the optimization function and determine logarithmic shadow prices to assess limitations during process time. This approach, which is quite similar to the approach described in this thesis (Hofer, Hauer, et al. 2018), is more applicable and mitigates many of the shortcomings of the publications of Xie and Wang (1994b) and Xie and Wang (1994a) and Xing et al. (2011).

Similar research has also been done on *Penicillium chrysogenum* in the past. An overview of published MFA models for this organism is given in table 1.1. Still, while these models in table 1.1 all incorporated the metabolism of amino acids, their focus was not explicitly on the elucidation of the effect of corn steep liquor. Furthermore, the few research groups that have used MFA for exploring *Penicillium chrysogenum* have shifted from pure metabolic flux balancing to mostly isotope labelled MFA. While from an academic point of view this extends the possibilities to take an ever more detailed look into the metabolism of this important biotechnological organism, from an industrial viewpoint this might

limit and hinder the application of this technique in a manufacturing setting. The focus on ^{13}C-labelled substrates means that the created models can never be applied in a manufacturing setting for process control and quality assurance.

References

Publications of special significance and value have their titles highlighted in boldface and have been complemented by an explanatory comment.

Abraham, E. P. and E. Chain (1940). "An enzyme from bacteria able to destroy penicillin". In: *Reviews of infectious diseases* 10.4, p. 677.

Aiba, S. and M. Matsuoka (1979). "**Identification of metabolic model: citrate production from glucose by *Candida lipolytica***". In: *Biotechnology and Bioengineering* 21.8, pp. 1373–1386. DOI: 10.1002/bit.260210806.
According to Çalık & Özdamar (2002) this publication constitutes the origins of MFA. Stephanopoulos, Aristidou & Nielsen (1998) also believe this to *probably* be the first time someone used metabolite balancing on fermentation data.

Antoniewicz, M. R. (2015). "Methods and advances in metabolic flux analysis: a mini-review". In: *Journal of industrial microbiology & biotechnology* 42.3, pp. 317–325. DOI: 10.1007/s10295-015-1585-x.

Beg, Q. K., V. Sahai, and R. Gupta (2003). "Statistical media optimization and alkaline protease production from *Bacillus mojavensis* in a bioreactor". In: *Process Biochemistry* 39.2, pp. 203–209. DOI: 10.1016/S0032-9592(03)00064-5.

Bickel, L. (1972). *Rise Up to Life*. London: Angus and Robertson.

Böhm, J. et al. (2013). "Sexual reproduction and mating-type–mediated strain development in the penicillin-producing fungus *Penicillium chrysogenum*". In: *Proceedings of the National Academy of Sciences* 110.4, pp. 1476–1481. DOI: 10.1073/pnas.1217943110.

Bycroft, B. W. and R. E. Shute (1987). "Chemistry and biosynthesis of penicillins and cephalosporins". In: *Penicillium and Acremonium*. Springer, pp. 113–160. DOI: 10.1007/978-1-4899-1986-1_5.

Calam, C. T. and B. A. Ismail (1980). "Investigation of factors in the optimisation of penicillin production". In: *Journal of Chemical Technology and Biotechnology* 30.1, pp. 249–262. DOI: 10.1002/jctb.503300131.

Çalık, P. and T. H. Özdamar (2002). "Bioreaction network flux analysis for industrial microorganisms: A review". In: *Reviews in Chemical Engineering* 18.6, pp. 553–604. DOI: 10.1515/REVCE.2002.18.6.553.

Caspi, R. et al. (2006). "MetaCyc: a multiorganism database of metabolic pathways and enzymes". In: *Nucleic acids research* 34.suppl1, pp. D511–D516. DOI: 10.1093/nar/gkj128.

Desai, K. M. et al. (2008). "Comparison of artificial neural network (ANN) and response surface methodology (RSM) in fermentation media optimization: case study of fermentative production of scleroglucan". In: *Biochemical Engineering Journal* 41.3, pp. 266–273. DOI: 10.1016/j.bej.2008.05.009.

Edwards, A. G. and C. S. Ho (1988). "Effects of carbon dioxide on *Penicillium chrysogenum*: an autoradiographic study". In: *Biotechnology and bioengineering* 32.1, pp. 1–7. DOI: 10.1002/bit.260320103.

Feist, A. M. and B. Ø. Palsson (2010). "The biomass objective function". In: *Current opinion in microbiology* 13.3, pp. 344–349. DOI: 10.1016/j.mib.2010.03.003.

Fernandes, P. and E. Martens (2017). "Antibiotics in late clinical development". In: *Biochemical pharmacology* 133, pp. 152–163. DOI: 10.1016/j.bcp.2016.09.025.

Finch, R. G. et al. (2010). *Antibiotic and chemotherapy e-book*. Elsevier Health Sciences.

Fleming, A. (1929). "**On the antibacterial action of cultures of a *Penicillium*, with special reference to their use in the isolation of *B. influenzæ***". In: *Br J Exp Pathol* 10, pp. 226–236.
This is the first publication in which Fleming describes his discovery of penicillin.

Galleguillos, S. N. et al. (2017). "What can mathematical modelling say about CHO metabolism and protein glycosylation?" In: *Computational and structural biotechnology journal* 15, pp. 212–221. DOI: doi.org/10.1016/j.csbj.2017.01.005.

Gheshlaghi, R. E. Z. A. et al. (2007). "Metabolic flux analysis for optimizing the specific growth rate of recombinant *Aspergillus niger*". In: *Bioprocess and biosystems engineering* 30.6, pp. 397–418. DOI: 10.1007/s00449-007-0136-x.

Gouda, M. K., A. E. Swellam, and S. H. Omar (2001). "Production of PHB by a *Bacillus megaterium* strain using sugarcane molasses and corn steep liquor as sole carbon and nitrogen sources". In: *Microbiological research* 156.3, pp. 201–207. DOI: 10.1078/0944-5013-00104.

Gudiña, E. J. et al. (2015). "Biosurfactant production by *Bacillus subtilis* using corn steep liquor as culture medium". In: *Frontiers in microbiology* 6, p. 59. DOI: 10.3389/fmicb.2015.00059.

Hare, R. (1982). "New light on the history of penicillin". In: *Medical history* 26.1, pp. 1–24. DOI: 10.1017/S00257273000407580.

Harris, L. M. et al. (2000). "Characterization of recombinant strains of the *Clostridium acetobutylicum* butyrate kinase inactivation mutant: need for new phenomenological models for solventogenesis and butanol inhibition?" In: *Biotechnology and bioengineering* 67.1, pp. 1–11. DOI: 10.1002/(SICI)1097-0290(20000105)67:1%3C1::AID-BIT1%3E3.0.CO;2-G.

Henriksen, C. M. et al. (1996). "Growth energetics and metabolic fluxes in continuous cultures of *Penicillium chrysogenum*". In: *Journal of Biotechnology* 45.2, pp. 149–164. DOI: 10.1016/0168-1656(95)00164-6.

Hofer, A., S. Hauer, et al. (2018). "Metabolic flux analysis linked to complex raw materials as tool for bioprocess improvement". In: *Chemical Engineering Science* 191, pp. 245–252. DOI: 10.1016/j.ces.2018.06.075.

Hofer, A. and C. Herwig (2017). "Quantitative determination of nine water-soluble vitamins in the complex matrix of corn steep liquor for raw material quality assessment". In: *Journal of Chemical Technology & Biotechnology* 92.8, pp. 2106–2113. DOI: 10.1002/jctb.5211.

Houbraken, J., J. C. Frisvad, and R. A. Samson (2011). "Fleming's penicillin producing strain is not *Penicillium chrysogenum* but *P. rubens*". In: *IMA fungus* 2.1, pp. 87–95. DOI: 10.5598/imafungus.2011.02.01.12.

Hull, S. R. et al. (1996). "Composition of corn steep water during steeping". In: *Journal of agricultural and food chemistry* 44.7, pp. 1857–1863. DOI: 10.1021/jf950353v.

Jami, M.-S. et al. (2010). "Proteome analysis of the penicillin producer *Penicillium chrysogenum*: characterization of protein changes during the industrial strain improvement". In: *Molecular & Cellular Proteomics* 9.6, pp. 1182–1198. DOI: 10.1074/mcp.M900327-MCP200.

Jørgensen, H. S., H. Møllgaard, et al. (1992). "Identification of rate controlling enzymes in the biosynthetic pathway to penicillin in a high yielding strain of *Penicillium chrysogenum*". In: *IFAC Proceedings Volumes* 25.2, pp. 251–254. DOI: 10.1016/S1474-6670(17)50364-X.

Jørgensen, H. S., J. Nielsen, et al. (1995). "Metabolic flux distributions in *Penicillium chrysogenum* during fed-batch cultivations". In: *Biotechnology and Bioengineering* 46.2, pp. 117–131. DOI: 10.1002/bit.260460205.

Jüsten, P. et al. (1998). "Dependence of *Penicillium chrysogenum* growth, morphology, vacuolation, and productivity in fed-batch fermentations on impeller type and agitation intensity". In: *Biotechnology and bioengineering* 59.6, pp. 762–775. DOI: 10.1002/(SICI)1097-0290(19980920)59:6%3C762::AID-BIT13%3E3.0.CO;2-7.

Kiel, J. A. K. W. et al. (2005). "Overproduction of a single protein, Pc-Pex11p, results in 2-fold enhanced penicillin production by *Penicillium chrysogenum*". In: *Fungal Genetics and Biology* 42.2, pp. 154–164. DOI: 10.1016/j.fgb.2004.10.010.

Kona, R. P., N. Qureshi, and J. S. Pai (2001). "Production of glucose oxidase using *Aspergillus niger* and corn steep liquor". In: *Bioresource technology* 78.2, pp. 123–126. DOI: 10.1016/S0960-8524(01)00014-1.

Lee, P. C. et al. (2000). "Fermentative production of succinic acid from glucose and corn steep

liquor by *Anaerobiospirillum succiniciproducens*". In: *Biotechnology and bioprocess engineering* 5.5, pp. 379–381. DOI: 10.1007/BF02942216.

Ligon, B. L. (2004). "Penicillin: its discovery and early development". In: *Seminars in pediatric infectious diseases*. Vol. 15. 1. Elsevier, pp. 52–57. DOI: 10.1053/j.spid.2004.02.001.

Liu, X. et al. (2015). "Citric acid production by *Yarrowia lipolytica* SWJ-1b using corn steep liquor as a source of organic nitrogen and vitamins". In: *Industrial Crops and Products* 78, pp. 154–160. DOI: 10.1016/j.indcrop.2015.10.029.

Long, C. P. and M. R. Antoniewicz (2014). "Metabolic flux analysis of *Escherichia coli* knockouts: lessons from the Keio collection and future outlook". In: *Current opinion in biotechnology* 28, pp. 127–133. DOI: 10.1016/j.copbio.2014.02.006.

Makagiansar, H. Y. et al. (1993). "The influence of mechanical forces on the morphology and penicillin production of *Penicillium chrysogenum*". In: *Bioprocess Engineering* 9.2-3, pp. 83–90. DOI: 10.1007/BF00369035.

Mantle, P. G. (1987). "Secondary metabolites of *Penicillium* and *Acremonium*". In: *Penicillium and Acremonium*. Springer, pp. 161–243. DOI: 10.1007/978-1-4899-1986-1_6.

McIntyre, M., D. R. Berry, and B. McNeil (2000). "Role of proteases in autolysis of *Penicillium chrysogenum* chemostat cultures in response to nutrient depletion". In: *Applied microbiology and biotechnology* 53.2, pp. 235–242. DOI: 10.1007/s002530050014.

Miles, E. A. and A. P. J. Trinci (1983). "Effect of pH and temperature on morphology of batch and chemostat cultures of *Penicillium chrysogenum*". In: *Transactions of the British Mycological Society* 81.2, pp. 193–200. DOI: 10.1016/S0007-1536(83)80069-2.

Miller, E. L. (2002). "The penicillins: a review and update". In: *Journal of midwifery & women's health* 47.6, pp. 426–434. DOI: 10.1016/S1526-9523(02)00330-6.

Moss, M. O. (1987). "Morphology and physiology of *Penicillium* and *Acremonium*". In: *Penicillium and Acremonium*. Springer, pp. 37–71. DOI: 10.1007/978-1-4899-1986-1_2.

Moyer, Andrew J. and Robert D. Coghill (1946). "Penicillin: VIII. Production of Penicillin in Surface Cultures 1". In: *Journal of bacteriology* 51.1, p. 57.

Newbert, R. W. et al. (1997). "Analysis of a commercially improved *Penicillium chrysogenum* strain series: involvement of recombinogenic regions in amplification and deletion of the penicillin biosynthesis gene cluster". In: *Journal of Industrial Microbiology and Biotechnology* 19.1, pp. 18–27. DOI: 10.1038/sj.jim.2900411.

Nielsen, J. et al. (1995). "Pellet formation and fragmentation in submerged cultures of *Penicillium chrysogenum* and its relation to penicillin production". In: *Biotechnology Progress* 11.1, pp. 93–98. DOI: 10.1021/bp00031a013.

Niklas, J. et al. (2011). "Quantitative characterization of metabolism and metabolic shifts during growth of the new human cell line AGE1. HN using time resolved metabolic flux analysis". In: *Bioprocess and biosystems engineering* 34.5, pp. 533–545. DOI: 10.1007/s00449-010-0502-y.

Nissen, T. L. et al. (1997). "Flux distributions in anaerobic, glucose-limited continuous cultures of *Saccharomyces cerevisiae*". In: *Microbiology* 143.1, pp. 203–218. DOI: 10.1099/00221287-143-1-203.

Nolan, R. P. and K. Lee (2011). "Dynamic model of CHO cell metabolism". In: *Metabolic engineering* 13.1, pp. 108–124. DOI: 10.1016/j.ymben.2010.09.003.

O'Neill, J. (2016). *Review on Antimicrobial Resistance. Tackling drug-resistant infections globally: Final report and recommendations. May 2016*.

Ogata, H. et al. (1999). "KEGG: Kyoto encyclopedia of genes and genomes". In: *Nucleic acids research* 27.1, pp. 29–34.

Ohadi, K., R. L. Legge, and H. M. Budman (2015). "Development of a soft-sensor based on multi-wavelength fluorescence spectroscopy and a dynamic metabolic model for monitoring mammalian cell cultures". In: *Biotechnology and bioengineering* 112.1, pp. 197–208. DOI: 10.1002/bit.25339.

Papoutsakis, E. T. (1984). "Equations and calculations for fermentations of butyric acid bacteria". In: *Biotechnology and bioengineering* 26.2, pp. 174–

187. DOI: 10.1002/(SICI)1097-0290(20000320)67:6%3C813::AID-BIT17%3E3.0.CO;2-X.

Papoutsakis, E. T. and C. L. Meyer (1985). "Equations and calculations of product yields and preferred pathways for butanediol and mixed-acid fermentations". In: *Biotechnology and bioengineering* 27.1, pp. 50–66. DOI: 10.1002/bit.260270108.

Price, K. E. (1969). "Structure-activity relationships of semisynthetic penicillins". In: *Adv. Appl. Microbiol* 11.17-75, p. 930. DOI: 10.1016/S0065-2164(08)70606-3.

Prosser, J. I. (1995). "Kinetics of filamentous growth and branching". In: *The growing fungus*. Springer, pp. 301–318. DOI: 10.1007/978-0-585-27576-5_14.

Provost, A. and G. Bastin (2004). "**Dynamic metabolic modelling under the balanced growth condition**". In: *Journal of Process Control* 14.7, pp. 717–728. DOI: 10.1016/j.jprocont.2003.12.004.
A detailed description of the general dynamic metabolic model, metabolic flux analysis, and how to generate macro-reactions from the stoichiometric model. Uses simple Monod kinetics for dynamic modelling of the macro-reactions, which were dereived from pathway analysis by using elementary flux modes. Based on the assumptions of balanced and exponential growth.

El-Sabbagh, N., B. McNeil, and L. M. Harvey (2006). "Dissolved carbon dioxide effects on growth, nutrient consumption, penicillin synthesis and morphology in batch cultures of *Penicillium chrysogenum*". In: *Enzyme and microbial technology* 39.2, pp. 185–190. DOI: doi.org/10.1016/j.enzmictec.2005.10.020.

Salo, O. V. et al. (2015). "Genomic mutational analysis of the impact of the classical strain improvement program on β–lactam producing *Penicillium chrysogenum*". In: *BMC genomics* 16.1, p. 937. DOI: 10.1186/s12864-015-2154-4.

Schmidt, L. H. and C. L. Sesler (1943). "Development of resistance to penicillin by pneumococci". In: *Proceedings of the Society for Experimental Biology and Medicine* 52.4, pp. 353–357. DOI: 10.3181%2F00379727-52-14153.

Schnitzer, R. J., L. J. Camagni, and M. Buck (1943). "Resistance of small colony variants (G-forms) of a *Staphylococcus* towards the bacteriostatic activity of penicillin". In: *Proceedings of the Society for Experimental Biology and Medicine* 53.1, pp. 75–78. DOI: 10.3181%2F00379727-53-14192.

Schwender, J. (2008). "Metabolic flux analysis as a tool in metabolic engineering of plants". In: *Current Opinion in Biotechnology* 19.2, pp. 131–137. DOI: 10.1016/j.copbio.2008.02.006.

Scott, J. et al. (2004). "Genotypic variation in *Penicillium chysogenum* from indoor environments". In: *Mycologia* 96.5, pp. 1095–1105. DOI: 10.1080/15572536.2005.11832908.

Silveira, M. et al. (2001). "Production of glucose–fructose oxidoreductase and ethanol by *Zymomonas mobilis* ATCC 29191 in medium containing corn steep liquor as a source of vitamins". In: *Applied microbiology and biotechnology* 55.4, pp. 442–445. DOI: 10.1007/s002530000569.

Singh, V. et al. (2017). "Strategies for fermentation medium optimization: an in-depth review". In: *Frontiers in microbiology* 7, p. 2087. DOI: 10.3389/fmicb.2016.02087.

Smith, J. J., M. D. Lilly, and R. I. Fox (1990). "The effect of agitation on the morphology and penicillin production of *Penicillium chrysogenum*". In: *Biotechnology and bioengineering* 35.10, pp. 1011–1023. DOI: 10.1002/bit.260351009.

Stephanopoulos, G., A. A. Aristidou, and J. Nielsen (1998). ***Metabolic Engineering: principles and methodologies***. ISBN: 9780080536286.
The value of this book cannot be overstated. Chapters 3 and 8 of this educational book give a very good overview of the theory and applications behind MFA. This volume should be considered the go-to work if one wishes to accustume themselves to MFA, MCA and metabolic engineering in general.

Stephanopoulos, G. and J. J. Vallino (1991). "Network rigidity and metabolic engineering in metabolite overproduction". In: *Science* 252.5013, pp. 1675–1681. DOI: 10.1126/science.1904627.

Van Gulik, W. M. et al. (2000). "Application of metabolic flux analysis for the identification of metabolic bottlenecks in the biosynthesis of penicillin-G". In: *Biotechnology and Bioengineering* 68.6, pp. 602–618. DOI: 10.1002/(SICI)1097-

0290(20000620)68:6%3C602::AID-BIT3%3E3.0.CO;2-2.

Varma, A. and B. Ø. Palsson (1994). "Metabolic flux balancing: basic concepts, scientific and practical use". In: *Bio/technology* 12.10, pp. 994–998. DOI: 10.1038/nbt1094-994.

Verhoff, F. H. and J. E. Spradlin (1976). "Mass and energy balance analysis of metabolic pathways applied to citric acid production by *Aspergillus niger*". In: *Biotechnology and Bioengineering* 18.3, pp. 425–432. DOI: 10.1002/bit.260180312.

Waxman, D. J. and J. L. Strominger (1983). "Penicillin-binding proteins and the mechanism of action of beta-lactam antibiotics". In: *Annual review of biochemistry* 52.1, pp. 825–869. DOI: 10.1146/annurev.bi.52.070183.004141.

Weber, S. S. et al. (2012). "Increased penicillin production in *Penicillium chrysogenum* production strains via balanced overexpression of isopenicillin N acyltransferase". In: *Appl. Environ. Microbiol.* 78.19, pp. 7107–7113. DOI: 10.1128/AEM.01529-12.

Wiechert, W. (2001). "^{13}C metabolic flux analysis". In: *Metabolic engineering* 3.3, pp. 195–206. DOI: 10.1006/mben.2001.0187.

Winden, W. van, P. Verheijen, and S. Heijnen (2001). "Possible pitfalls of flux calculations based on ^{13}C-labeling". In: *Metabolic engineering* 3.2, pp. 151–162. DOI: 10.1006/mben.2000.0174.

Xie, L. and D. I. C. Wang (1994a). "Fed-batch cultivation of animal cells using different medium design concepts and feeding strategies". In: *Biotechnology and bioengineering* 43.11, pp. 1175–1189. DOI: 10.1002/bit.260431123.

– (1994b). "Stoichiometric analysis of animal cell growth and its application in medium design". In: *Biotechnology and bioengineering* 43.11, pp. 1164–1174. DOI: 10.1002/bit.260431122.

Xing, Z. et al. (2011). "Optimizing amino acid composition of CHO cell culture media for a fusion protein production". In: *Process Biochemistry* 46.7, pp. 1423–1429. DOI: 10.1016/j.procbio.2011.03.014.

Zervosen, A. et al. (2012). "Development of new drugs for an old target — the penicillin binding proteins". In: *Molecules* 17.11, pp. 12478–12505. DOI: 10.3390/molecules171112478.

Zhang, J. et al. (2003). "Toward consistent and productive complex media for industrial fermentations: studies on yeast extract for a recombinant yeast fermentation process". In: *Biotechnology and bioengineering* 82.6, pp. 640–652. DOI: 10.1002/bit.10608.

Chapter 2

Goal of the Thesis

> [...] modeling is relatively meaningless without explicit definition, at the outset, of its purpose.
>
> James E. Bailey

SIMPLY fitting data points through some lines may serve as an intellectual exercise in solving puzzles but will not push back the frontiers of science. As a consequence, every model needs a clearly defined problem or a very specific goal and purpose. This does not only help in communicating the ideas behind the model and getting other people at all interested in your research, but also aids in not losing track of what is important.

Scientific questions

THE specific goal of this thesis and the model described within, was to develop a method with which the effect of media components can be displayed over time. It was applied to and verified by fermentations of penicillin by *Penicillium chrysogenum* on glucose with added corn steep liquor as complex media additive. The focus of interest were the amino acids within the corn steep liquor as penicillin itself is a peptide. Within this defined goal, this leads to the formulation of a few key scientific questions:

- What are the release kinetics of amino acids from corn steep liquor?
- What is the role of amino acids on penicillin formation? Which are limiting over process time and why?
- What is the role of amino acids on biomass formation? Which are limiting over process time and why?
- How do energetic demands guide and limit penicillin formation?
- Do metabolic shifts occur during fermentation? Can these shifts be used for process characterization?

Quality by design

HOWEVER, the answers to these questions should not be evaluated on their own, but in the wider context of quality by design (QbD). In the last decades there has been a paradigm shift by regulatory bodies regarding the definition of quality, which is best exemplified by the International Council for Harmonisation (ICH) guideline Q8 (R2): "It is important to recognize that quality cannot be tested into products; i.e., quality should be built in by design." This in turn, has led to a definition of and focus on *Critical Process Parameters* and *Critical Material Attributes* to facilitate this transition from *product testing* to *process control* as a means of achieving quality. The scientific questions answered therefore have themselves a purpose in giving a knowledge and data-based rationale for

defining critical material attributes that affect process performance.

Chapter 3

Methodical Basis

> The study of the growth of bacterial cultures does not constitute a specialised subject or a branch of research: it is the basic method of microbiology.
>
> Jacques Monod

THIS section is written as to give a basic overview of the most important aspects of this thesis. It is supposed to act not only as a starting point, and a collection of the relevant literature, but also as a short and concise instruction for anyone that wants to learn or refresh their knowledge about these topics. This section was written in a way so that it can be handed to a student and used as easy reference, explaining difficult concepts and eliminating common misconceptions that otherwise would lead to much frustration on the student's part. In essence I wanted to collect all knowledge that he wished he had possessed at the beginning of his thesis. As a consequence, this section might be a bit more elaborate than expected or deemed necessary, however, it is hoped that somebody someday finds this helpful.

Determination of reaction rates and kinetics

IN biotechnological processes most analytical measurements quantify concentrations of substrates and products in the fermentation broth. However, these measurement results convey little biologically relevant information by themselves, so the first step in gaining insights from fermentation data is the calculation of specific uptake and production rates q_i for each substrate and product i, respectively. This can be done by the definition of ordinary differential equations (ODE) and subsequent numerical solution. For a typical fermentation process these ODEs form a material balance, where the change in concentration c_i depends on the specific uptake or production rate q_i, the amount of biomass X, and the volume flows in and out of the reactor $\dot{V}_{in}$ and $\dot{V}_{out}$.

$$\frac{\delta c_i}{\delta t} = q_i X + \dot{V}_{in} c_{i,in} - \dot{V}_{out} c_i \tag{3.1}$$

In the simplest case each q_i is assumed to be constant for each interval between measurements or for a phase defined over several measurements (Kroll et al. 2017). Alternatively, q_i is often expected to follow Monod-kinetics (Rickenberg et al. 1956) in which case the numerical solution of the ODE will return the kinetic parameters that seem to best describe the process. For a biochemical reaction r_j that turns n_S substrates S_l into n_P products P_m with the stoichiometric factors α_l and β_m respectively,

$$\sum_{l=1}^{n_S} \alpha_l S_l \xrightarrow{r_j} \sum_{m=1}^{n_P} \beta_m P_m \tag{3.2}$$

it can be assumed that more than one substrate is limiting and that there exist interaction effects be-

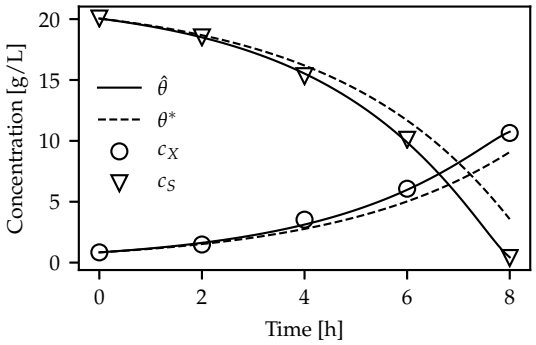

Figure 3.1: Biomass and substrate concentrations c_X and c_S against time for a simple batch fermentation following Monod kinetics. Measured data points are signified by circles for c_X and triangles for c_S. It is important to treat the initial conditions also as parameters, or else the estimates shown as lines will be misleading, even though the resulting calculated concentrations seem to fit the data well. This is exemplified by comparing the concentrations calculated by the estimated parameters $\hat{\theta}$ (solid line)and the true parameters θ^* (dashed line), when ignoring the initial conditions as parameters.

tween substrate limitations. In this case, the uptake/production rate will follow the form

$$q_i = q_{j,max}\alpha_i \frac{\prod_{l=1}^{n_S} c_{S_l}^{\alpha_l}}{\prod_{l=1}^{n} (k_l + c_{S_l})}, \tag{3.3}$$

When fitting data in this way it should be emphasized, that for the solution of ODEs, the initial concentration at $t = 0$ plays an important role and will greatly influence the results of parameter fitting. However, the true concentration at this point in time is not exactly known, and the measurements themselves are subject to random error. It would be a grave mistake to take the measured concentrations at $t = 0$ as the initial conditions as this would lead to a wrong estimation of parameters as is demonstrated in figure 3.1. The initial concentrations should therefore be treated just as parameters and be estimated by least-squares methods (Gujer 2008).

It should be emphasized that while Michaelis-Menten and Monod-kinetics have become the de-facto standard models in many fermentation processes, they are not the only and not necessarily the best approach. While their popularity is due to its simplicity and elegance, it has been known for a very long time that they do not necessarily fit experimental data particularly well (Powell 1967; Dabes, Finn, and Wilke 1973; Bader 1978). Koch (1998) has even gone so far as calling the Monod-model as being "theoretically and actually wrong in almost all cases!" Alternatives models for growth rate would be Blackman (1905), exponential growth, Contois (1959), or a three-constant model that is a generalisation of Blackmann and Monod (Dabes, Finn, and Wilke 1973). Artificial neural networks have also been successfully applied to this problem and seem to be the best current method for exact results (Basheer and Hajmeer 2000). However, as a black-box model it is a bit unsatisfactory as the parameters of mechanistic models give us meaningful ways of classifying and comparing strains, and gives us insights on which parameters need to be improved to increase rates, which ANN do not provide. Table 3.1 is supposed to give an overview of a few simple mechanistic models and their basic assumptions.

However, caution should be applied for any of these simple approximations, as the "constants" of these models are themselves highly variable based on growth conditions and physiological state of the cell, which in turn are influenced by factors like genotype, inoculum history, cell density and starvation length (Ferenci 1999). Kovárová-Kovar and Egli (1998) bemoan that despite the impression that some textbooks may give on growth rates (and in conclusion also on substrate uptake and production rates) "many fundamental questions in this field are still waiting to be discovered, established, and exploited".

Parameter identifiability analysis

For parametric models as for instance the Monod-model, one often determines the parameters by fitting them to experimentally determined data. Purely visual inspection between predicted and observed values alone, while being absolutely indispensable is not sufficient. A good fit only indicates

Table 3.1: Overview of simple mechanistic models that can be applied to the description of the growth rate.

Model	Form
Monod (1941)	$q_{max}\frac{c_S}{K+c_S}$
Blackman (1905)	$\begin{cases} Kc_S & \text{if} \quad S < \frac{q_{max}}{K} \\ q_{max} & \text{else} \end{cases}$
Best (1955)	$\frac{q_{max}}{2J}(c_S + K + J)\left(1 - \sqrt{1 - \frac{4c_S J}{(c_S+K+J)^2}}\right)$
Contois (1959)	$q_{max}\frac{c_S}{Kc_X+c_S}$
Heijnen and Romein (1995)	$q_{max}\left[\frac{c_S}{c_S-1+2^{1/n}}\right]^n$

that the model structure and parameters are able to describe the data well but it does not mean that there is a *unique* set of parameters for the specific data set. Identifiability analysis concerns itself exactly with this question (Walter 2014) and therefore can be used to detect overfitting due to an inappropriate model or insufficient data (Raue, Kreutz, et al. 2009; Raue, Becker, et al. 2010). Several independent approaches exist to identifiability analysis such as *differential algebra identifiability of systems* (Bellu et al. 2007) or *exact arithmetic rank* (Sedoglavic 2002). Here only the approach utilizing profile likelihoods will be described as this approach is able to also detect practical non-identifiability and can be applied to models using non-rational expressions like if-statements (Raue, Karlsson, et al. 2014).

Fitting of a model to data is done by minimising an objective function. Often this will be the sum of squared residuals, where the residuals can be normalised to the error on the measured variable.

$$\chi^2(\theta) = \sum_{k=1}^{m}\sum_{l=1}^{d}\left(\frac{y_{kl}^D - y_k(\theta, t_l)}{\sigma_{kl}^D}\right)^2 \tag{3.4}$$

The profile likelihood χ^2_{PL} can be estimated by this optimisation function. This is done for each parameter θ_i, by setting it to a defined value and then fitting the other parameters to the data with this specified θ_i. When doing this over a wide range, we get the minimal possible sum of squared residuals as a function of θ_i, which can be used as a placeholder for the likelihood (Raue, Kreutz, et al. 2009).

$$\chi^2_{PL}(\theta_i) = \min_{\theta_{j\neq i}}[\chi^2(\theta)] \tag{3.5}$$

If we now assume that the difference $\chi^2(\theta^*) - \chi^2(\hat{\theta})$ follows the χ^2 distribution, where θ^* are the true and $\hat{\theta}$ the predicted parameters, we can determine confidence intervals of each parameter θ_i based on the profile likelihood. The confidence intervals are defined as where the profile likelihood exceeds the value of the χ^2 distribution for a given α and degrees of freedom.

$$\Delta_\alpha = \chi^2(\alpha, df) \tag{3.6}$$

Raue, Kreutz, et al. (2009) use $\alpha = 0.68$. The degrees of freedom can be either set to 1 for the *pointwise confidence interval* that holds for each parameter individually or at the number of parameters for *simultaneous confidence intervals* that hold true for all parameters. The first is considered less reliable by Raue, Kreutz, et al. (2009). It should also be noted that the confidence intervals are based on the assumption that the difference $\chi^2(\theta^*) - \chi^2(\hat{\theta})$ follows the χ^2 distribution. To be certain that this holds true, the distribution of the difference "should always be checked by simulation" and δ_α be adjusted to the in this way observed distribution (Raue, Becker, et al. 2010). Alternative distributions to χ^2 for Δ might be the inverse-Gaussian, Weibull, or log-logistic distribution (Dick 2004).

A parameter is said to be structural non-identifiable if the confidence interval is infinite and

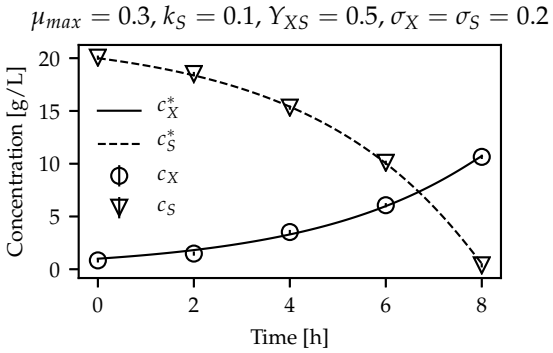

Figure 3.2: Example of a simulated batch fermentation with the true concentrations that follow Monod-kinetics as lines (c_X^* as a solid and c_S^* as a dashed line), and the simulated measured concentrations as markers (c_X as circles and c_S as triangles).

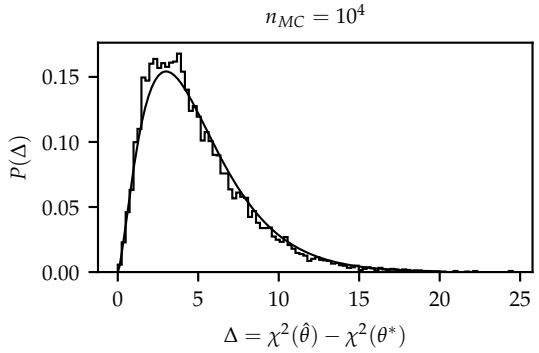

Figure 3.3: The difference in the sums of squared residuals between the estimated parameters with the best fit $\hat{\theta}$ and the true parameters θ follow a χ^2 distribution with five degrees of freedom for the five fitted parameters μ_{max}, k_S, Y_{XS}, $c_{X,0}$, and $c_{S,0}$. The empiric distribution was determined by Monte-Carlo simulation with $n_{MC} = 10^4$ samples drawn. The parameters in each sample are identical to the parameters described in the text or in figure 3.2.

there is no distinct minimum in the profile likelihood. Such a structural non-identifiability is due to the model structure itself and not due to imperfect data. It can be eliminated by simplifying the model, experimentally determining one or several parameters, or by measuring previously only estimated variables (Sedoglavic 2002).

A parameter is said to be practically non-identifiable if the confidence interval is infinite but there is a distinct minimum in the profile likelihood. Please note that a confidence interval is said to be infinite even if only one of its boarders is infinite. Practical non-identifiability arises due to suboptimal data, and can be eliminated by decreasing the error on measured variables, by increasing the sampling interval or changing the experimental conditions. If the confidence interval is finite and there is a minimum in the profile likelihood, then and only then a parameter is said to be identifiable.

A key strength of identifiability analysis by profile likelihoods is that each parameter can be determined completely on its own. Detaching the identifiability analysis between parameters leads to several advantages, including the possibility of parallelisation for faster computation, and easy visualisation and interpretation (Raue, Becker, et al. 2010). As the profile likelihood χ^2_{PL} as defined in equation 3.5 is not comparable between models, it can be normalised as described by Kroll et al. (2017) or by Meeker and Escobar (1995)

$$R(\chi^2) = \frac{\chi^2(\hat{\theta})}{\min\limits_{\theta_{j\neq i}}[\chi^2(\theta)]} \tag{3.7}$$

which changes the threshold for the confidence intervals to

$$\frac{\chi^2(\hat{\theta})}{\chi^2(\hat{\theta}) + \Delta_\alpha} \tag{3.8}$$

However, for both definitions of the profile likelihood this would mean that the threshold is not identical between experiments, parameters or models. Therefore, the profile likelihood will in the rest of this section be defined as

$$\chi^2_{PL}(\theta_i) = \min_{\theta_{j\neq i}}[\chi^2(\theta)] - \chi^2(\hat{\theta}) \tag{3.9}$$

so that the threshold value simply becomes Δ_{alpha}.

Let's apply this approach to a simple example: Assume that we have measured biomass and substrate concentrations c_X and c_S for a batch fermentation

at n_t time points and with known standard deviations of σ_X and σ_S, that are constant over time. Let's also assume that our data are perfectly described by a Monod-model where the true parameters are $\mu^*_{max} = 0.1$, $k^*_S = 0.01$, and $Y^*_{XS} = 0.5$ and the true initial conditions are $c^*_{X,0} = 1$, and $c^*_{S,0} = 20$. In figure 3.2 such a simulated experiment is shown, where $n_t = 5$ and $\sigma_X = \sigma_S = 0.2$ g/L.

By simulating $n_{MC} = 10^4$ samples and calculating the sum of squared residuals of the parameters with the best fit $\hat{\theta}$ and the true parameters θ^* we can make sure that their difference follows a χ^2 distribution as shown in figure 3.3. Therefore, we can calculate confidence intervals of the parameters through the 95$^{\text{th}}$ percentile of the χ^2 distribution.

If we take a look at the profile likelihood χ^2_{PL} for all three parameters in figure 3.4, we immediately see that one of them, k_S, is practically non-identifiable. The value of k_S can change between several orders of magnitude without affecting the fit to the data. So for these data, there is no unique solution to k_S. As the number of samples drawn has little effect on identifiability here (data not shown), if k_S is to be estimated, we would either need to reduce the error on measurements or change the experimental setup. Kovárová-Kovar and Egli (1998) believe that fed-batch fermentations are able to overcome many of the weaknesses of batch and continuous cultures in estimating growth parameters. However, it should be clearly emphasized that just because a parameter is identifiable this only means that there is a single parameter that gives a best fit to the data but not that this best fit parameter $\hat{\theta}_i$ is also close to the true parameter θ^*_i. This will be demonstrated by drawing not a single sample from the discussed system but one hundred and comparing the results of the identifiability analysis for k_S. In figure 3.5 A, we see that for $\sigma_X = \sigma_S = 0.2$ g/L many simulated experiments will be unable to identify k_S, however, many others will be able to identify it and will give results over several orders of magnitudes. Only by further decreasing the errors on measurements to 0.01, it is possible to estimate the true value of k_S reliably and consistently.

This just proves the importance of identifiability analysis in combination with extensive simulation for experimental planning.However, as the results of such a Monte-Carlo simulation depend also on the parameters of the system, this leads to the problematic conclusion that one needs to already know the parameters for simulation in order to be able to exactly determine the parameters experimentally. Hence, knowledge about the parameters is needed in order to plan the experiments for parameter determination.

Error propagation

Most properties of interest in the physical sciences cannot be measured directly but must be calculated or fitted from a combination of several other measurements. As such, on this result of these derived entities there must be an error that depends on the uncertainty of the measurements. Assignment of realistic errors for complex systems is not a trivial task. The most wide-spread taught way of estimating errors based on measurement uncertainty is Gaussian error propagation, which follows

$$\sigma_f^2 = \sum \left(\frac{\partial f(\theta)}{\partial \theta_i} \right)^2 \sigma_{\theta_i}^2 \tag{3.10}$$

However, equation 3.10 assumes that errors of variables are uncorrelated, which is often problematic, especially if parameters θ were themselves estimated by least-squares fit (Tellinghuisen 2001). Therefore, a more general and simpler matrix notation of equation 3.10 exists, in which correlation between variables is considered in the form of

$$\sigma_f^2 = g^T V g, \tag{3.11}$$

where V is the variance-covariance matrix and g^T is the sensitivity defined as the change in output per change in parameter value $g^T = (\frac{\partial f}{\partial \theta_1}, ..., \frac{\partial f}{\partial \theta_n})$. The sensitivity is calculated at each time point of a time series by approximating $\frac{\partial f}{\partial \theta_i}$ as $\frac{\Delta f}{\Delta \theta_i}$ using an appropriate value for $\Delta\theta$. Gujer (2008) proposes something in the range of one percent of the absolute parameter value. If too great a $\Delta\theta$ is chosen, one risks that the assumed linearity of the sensitivity does not

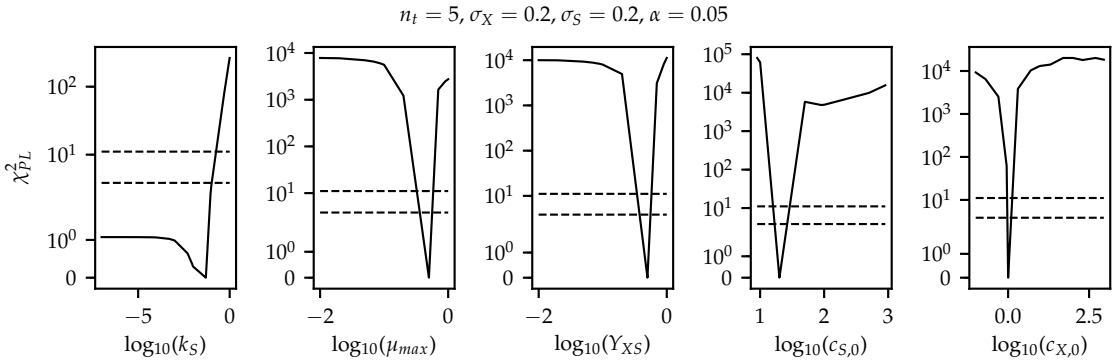

Figure 3.4: Result of the identifiability analysis of the simulated data from figure 3.2. $\Delta_{\alpha=0.05}$ both with df=1 and df=3 are indicated as dashed lines. While all other parameters are identifiable, k_S is practically non identifiable as its left confidence interval is $-\inf$ since it never rises above Δ_α.

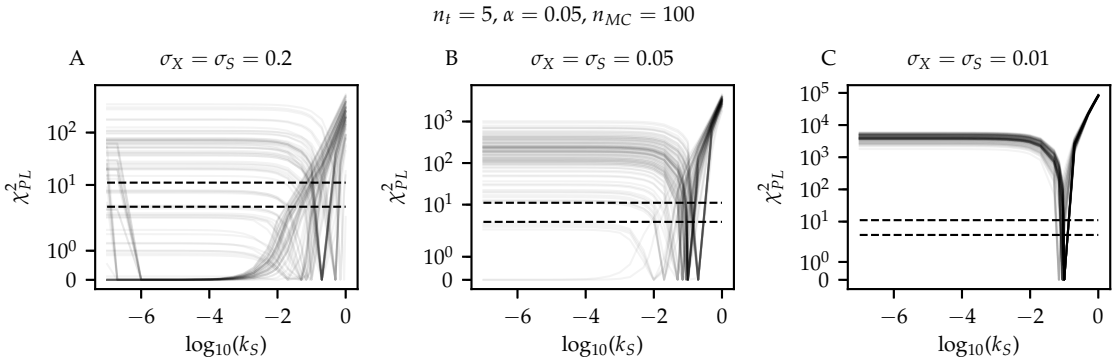

Figure 3.5: For three different true standard deviations (0.2 in A, 0.05 in B, and 0.01 in C) $n_{MC} = 100$ samples were drawn each and profile likelihoods were derived. The profile likelihoods for all samples were plotted slightly transparent so darker colours signify a more uniform distribution of profile likelihoods. With decreasing errors on measurements the reliability of estimation of k_S improves. For an error of 0.05, the estimate of k_S might still be several orders of magnitude away from the real parameter k_S^*, even though the identifiability analysis confirms that for this random sample there is a unique solution.

hold for non-linear models, which can lead to wrong estimates of uncertainty.

An alternative to Gaussian error propagation is the use of Monte-Carlo simulations. The method consists of three steps: The definition of input uncertainty, the sampling of input uncertainty and the calculation of outputs from sampled inputs to determine output uncertainty (Sin, Gernaey, and Lantz 2009). The strengths of Monte-Carlo simulation to the problem of error propagation are that it is easily implemented while being applicable to both linear and non-linear models and that it is able to deal with greater uncertainty than Gaussian error propagation (Gujer 2008). Furthermore, Monte-Carlo methods make no specific assumptions about the distributions of in and outputs, and the whole distribution of outputs can be assessed while linear error distribution only determines the momentum (Gujer 2008). Even though it is true that extensive Monte-Carlo simulations can be time consuming and computationally expensive, they are also ideally suited for parallel computing which can drastically decrease computation time. A bigger concern might be the properties of the random numbers generator (RNG) used for simulations. The desired properties of a good RNG are uniformity, independence, (and for cryptography also non-predictability) and to have these properties for a wide array of different sample sizes. All these properties can be checked by the use of freely available test suites. Failure of even one single test within such a test suite can already be problematic (L'Ecuyer and Simard 2007; O'Neill 2014). Press et al. (2007) give the following practical considerations for choosing a suitable RNG: i) Never use a linear congruent generator or multiplicative linear congruential generator. ii) Never use a generator with a period less than 2^{62}. They also advise against a period greater than 10^{100} as such a period should give no practical advantages. iii) Never use the built-in generators in C and C++. The RNG implemented in NumPy (v1.18) is PCG64 based on the publication by O'Neill (2014) and fulfills all modern requirements of RNGs. Additionally, when using parallel computing a RNG should be chosen that is able to provide multiple threads. Depending on the specific problem to which Monte-Carlo simulation is applied, Quasi-Monte-Carlo methods might be more effective and converging faster to the true solution, ironically even for reliability and uncertainty analysis (Robinson and Atcitty 1999; Singhee and Rutenbar 2010).

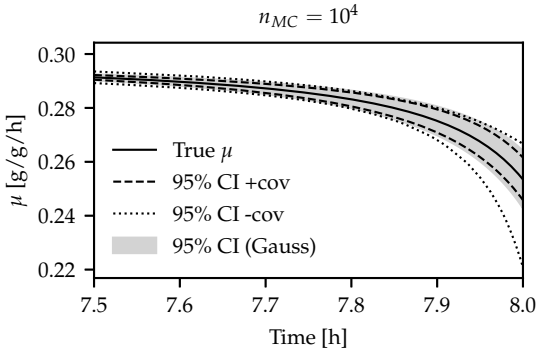

Figure 3.6: The 95 % confidence intervals calculated through Monte- Carlo simulations (dashed line) surround the true growth rate (solid line). When ignoring covariances between parameters (dotted line), the confidence interval becomes much wider. The 95 % confidence interval calculated through linear error propagation (including covariances) is depicted as an area in grey. Both Monte-Carlo and linear error propagation result in comparable boundaries.

For Monte-Carlo simulation, just as for linear error propagation, covariances between parameters need to be accounted for, or otherwise it could lead to an overestimation of output uncertainty (Gujer 2008). If for instance, the fitting of a Monod-model as described earlier would lead to an underestimation of $c_{X,0}$, then this could be compensated by an overestimation of μ_{max} to still lead to a good fit. It is therefore evident that when estimating parameters by least-squares, covariances should not be neglected. This is exemplified in figure 3.6, where ignoring covariances lead to wider confidence intervals for the growth rate μ. To incorporate covariances into the Monte-Carlo simulations of rates one has two approaches, either one simulates the error on measured data itself, or one generates the parameters from the estimations of mean and the variance-covariance matrix by assuming a multivariate normal distribution (Papoulis and Pillai 2002; Gujer 2008).

Metabolic flux analysis

> It should be emphpasized that metabolic fluxes are the most fundamental measure of cell physiology.
>
> Stephanopoulos, Aristidou and Nielsen in Metabolic Engineering

THE basis of MFA is a stoichiometric model describing how substrates S, products P, biomass constituents X_{Macro}, and metabolites X_{Met} are connected to each other through a number of reactions by their stoichiometric coefficients. This model, with stoichiometric factors for the four compound types grouped into four distinct stoichiometric matrices, can be formulated in by:

$$AS + BP + \Gamma X_{Macro} + GX_{Met} = 0 \tag{3.12}$$

The last of these four stoichiometric matrices are used for considering the properties of such a biological system by inspecting the mass balances of the intracellular metabolites. The change in concentrations within the cell itself is assumed by:

$$\frac{\delta c_{Met}}{\delta t} = r_{Met} - \mu c_{Met} \tag{3.13}$$

Under the assumptions of a pseudo-steady state and metabolite pools being several magnitudes smaller than their fluxes through them (Stephanopoulos, Aristidou, and Nielsen 1998), this simplifies to

$$0 = r_{Met} = G^T \times v \tag{3.14}$$

wherein G^T is the $n \times m$ stoichiometric matrix of the biochemical network consisting of n metabolites and m reactions, and v is the vector of m fluxes. This solution is sometimes also called the "right nullspace of the stoichiometric matrix" (Famili and Palsson 2003).

If the vector of measured fluxes v_m is sufficiently sized, so that the system is either determined or overdetermined, the vector of unknown fluxes v_c can be calculated after splitting the matrix G^T into the two matrices of unknown fluxes G_c^T and measured fluxes G_m^T, by the following equation:

$$G^T v = G_m^T v_m + G_c^T v_c = 0 \tag{3.15}$$

$$v_c = -(G_c^T)^{\#} G_m^T v_m \tag{3.16}$$

where

$$(G_c^T)^{\#} = (G_c G_c^T)^{-1} G_c \tag{3.17}$$

In this way, MFA can be understood as a form of least-squares linear regression (Jørgensen et al. 1995), that tries to minimize the residuals on the material balances. When solving overdetermined systems, the redundancy in measurements in combination with error estimates on the fluxes can be used to check the measurements on internal consistency (N. S. Wang and Stephanopoulos 1983).

$$R = G_m^T - G_c^T (G_c^T)^{\#} G_m^T \tag{3.18}$$

$$\epsilon = R_r v_m \tag{3.19}$$

$$F = E(\delta\delta^T) \tag{3.20}$$

$$\phi = R_r F R_r^T \tag{3.21}$$

$$h = \epsilon \phi^{-1} \epsilon^T \tag{3.22}$$

The calculated consistency index h follows a χ^2 distribution with the same degrees of freedom as the linear system. If the index h exceeds the threshold value at a given confidence interval, this could be due to gross measurement errors, violation of pseudo-steady state or an unsuitable model. For an overdetermined system gross measurement errors can be detected by iteratively eliminating a measurement and observing the index (Goudar et al. 2014). For testing if the pseudo-steady state is violated, one could iteratively add an extracellular output flux for each intracellular metabolite and again check for changes in consistency (Stephanopoulos, Aristidou, and Nielsen 1998). For internally consistent data,

Table 3.2: By iteratively removing a single reaction from the intracellular network G^T described by Goudar et al. (2014) it is demonstrated that the condition of a G_c^T is not bounded by the condition of G^T

Flux removed	None	v_{m1}	v_{m2}	v_{m3}	v_{m4}	v_{c1}	v_{c2}	v_{c3}	v_{c4}	v_{c5}	v_{c6}
Condition	7.6	20.0	8.8	36.5	29.3	8.4	7.7	17.3	11.3	81.1	7.71

improved measured rates can be estimated by the following formula:

$$\hat{v}_m = (I - F R_r^T \phi^{-1} R_r) v_m \tag{3.23}$$

However, before calculating intracellular fluxes in this manner, one should always consider the properties of the stoichiometric matrix first. One of these properties is the matrix rank, which should always be full, while the other is the condition number. The condition number is a measure of how errors in measurement and rounding effect the algebraic solution (Goudar et al. 2014; X. Wang 2007) and is bounded between 1 and $\inf$ (Allaire 2007). A system is said to be well-conditioned, if the condition number is below 100 while values above that suggest problems with sensitivity (Nyberg et al. 1999). As a rule of thumb it can be said that if the condition of A has the power b, that this is the upper bound of accuracy that one loses. In other words, this can be interpreted in the accuracy of the measurements needed for reliable results, e.g., if the power b would be 5, one would need to determine the measured fluxes to an accuracy of 5 digits (Stephanopoulos, Aristidou, and Nielsen 1998).

$$\kappa(A) = a \times 10^b \tag{3.24}$$

There seems to be some disagreement in the literature for which matrix exactly rank and condition should be observed. While Goudar et al. (2014) suggested it for G^T, in other publications as in Jørgensen et al. (1995) this is demanded for the network of unknown fluxes G_c^T. Stephanopoulos, Aristidou, and Nielsen (1998) demand a full rank of G_c^T but a low condition to G^T. To answer this question, we need to take a look at equation 3.17. For it to have a solution, we need the matrix $G_c G_c^T$ to be invertible, which is only the case if the columns of G_c^T are linearly independent, i.e., it is of full column rank (Strang 2005). Van der Heijden et al. (1994) have termed this property of the matrix as *non-redundant*. Stephanopoulos, Aristidou, and Nielsen (1998) also caution that a full rank of G does only mean that there exists at least one G_c^T of full rank but does not mean that every possible G_c^T has full rank. They fail to realise, however, that the same is true also for the condition of G^T and G_c^T, which is easily demonstrated by table 3.2.

However, in the case of rank deficiency one can still get results by calculating the pseudoinverse $(G_c^T)^{\#}$ by the Moore-Penrose method (Barata and Hussein 2012; Chen and Feng 2014), which is often also the method implemented in contemporary programming languages such as MatLab and numpy for Python. One can not stress enough that this is not a solution to a problem but only a dangerous liability. Therefore *both* G^T and G_c^T should always be checked for rank and condition. A more structural and comprehensive approach to this problem is given by calculability analysis and the interested reader is referred to the original publications by Van der Heijden et al. (1994) that have subsequently been built and improved upon, for instance by Klamt, S. Schuster, and Gilles (2002).

Describing a stoichiometric model

As already has been eluded to in the description of table 1.1, the reproduction of models described in the literature can be a little cumbersome at times. Some publications do not explicitly describe the reactions, while others do not describe the shape of the matrix and rarely is the distinction between substrates S, products P, macromolecular components X_{Macro} and intracellular metabolites X_{Met} stated

explicitly and in a comprehensive manner.. These things make reproducibility of MFA models a difficult process full of trial and error for the confused reader of publications. As it is clear that the inclusion of the matrices themselves in the publications is (usually) not in the best interest of publishers, authors and readers alike, the author would suggest the following minimum information to be included for every stoichiometric model:

- All intracellular reactions. Where compounds are included in these reactions for completeness without being part of the final modal, e.g., water, this should be indicated. The compound could for example either be put into round brackets or be crossed out.
- The compounds included in each of the four stoichiometric matrices α, B, Γ and G.
- Shape, rank, and condition of G^T. While the condition is meant to assure there are no problems of sensitivity, it can also serve as a practical way to confirm that one has correctly reproduced the matrix without errors.
- Shape, rank, and condition of G_c^T.

Creating a stoichiometric model

After, hopefully, having dispersed any confusion about matrix rank and condition, in the rest of this section important matrix properties will be discussed in detail with their effect on condition and flux calculation in mind and how to spot and deal with problems that may arise during model building.

Matrix size: At first, let us consider the case of a square $m \times m$ stoichiometric matrix G^T. As this system has m unknown fluxes and also m metabolic balances, its degrees of freedom at full rank is zero. In other words, the system that has been defined by this matrix is inherently determined. It follows that such a system is not able to adapt to changes in its environment as it itself is not capable of change. Therefore, it is evident that any stoichiometric matrix G^T must have more columns than rows.

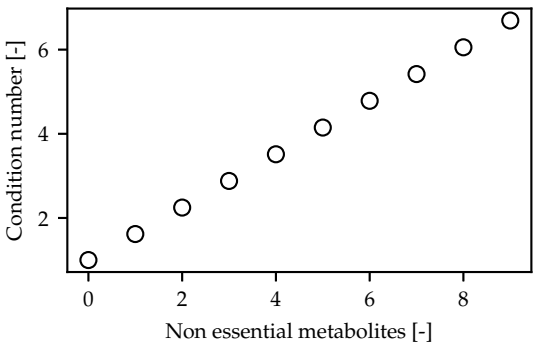

Figure 3.7: The condition number increases with the addition of nonbranching nodes.

Non-branching nodes: Lets consider the simplest possible branching network consisting of a substrate that is catalysed to either of two products.

$$S \xrightarrow{v_1} M \tag{3.25}$$

$$M \xrightarrow{v_2} P_1 \tag{3.26}$$

$$M \xrightarrow{v_3} P_2 \tag{3.27}$$

Such a system would consist of three fluxes, an uptake and two reaction fluxes and have the form the matrix G_0^T in equation 3.28. One might build on this network by adding n additional fluxes and nodes that serve no other purpose but to relay the input flux to the branching node. While such a modified model might be plausible from a biochemical point of view and might be used in metabolic control analysis (MCA) for simulation of perturbation experiments, from the standpoint of MFA this adds nothing of value since the input and output fluxes must be the same for each of these nodes due to the pseudo-steady state assumed. However, each of these added nodes will linearly increase the condition number, as is shown in figure 3.7.

$$G_0^T = \begin{bmatrix} 1 & -1 & -1 \end{bmatrix} \tag{3.28}$$

$$G_n^T = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots & \vdots \\ 0 & \dots & 1 & -1 & -1 \end{bmatrix} \tag{3.29}$$

Just as bad would it be, if one would formulate this stoichiometric matrix with the substrate S and the products P_1 and P_2 explicitly mentioned, as in equation 3.30.

$$G_0^T = \begin{bmatrix} -1 & 0 & 0 \\ 1 & -1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{3.30}$$

This would not only increase the condition from 1 to 2, but also be quite insensible as this would violate the pseudo-steady state assumption for these substances. By adding uptake fluxes to the matrix G^T to not violate the mass balances, we would again only create useless relay nodes. Therefore, each metabolite mass balance of the stoichiometric matrix should be checked to contain 3 or more non-zero stoichiometric factors. If a metabolite contains only one, it should be eliminated by removing the row, if it contains only two, the two input fluxes of the node should be combined into one. While these thoughts might be considered trivial, it is still important to mention them as these regrettable mistakes can be seen even in publications. A prime example comes from the model described in the otherwise excellent book chapter by Goudar et al. (2014) in Animal Cell Biotechnology. One can immediately see, that one can combine fluxes v_{m1} with v_{c1}, v_{m2} with v_{c2}, and v_{m3} with the two fluxes v_{c3} and v_{c5}.

$$G^T = \begin{bmatrix} -1 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 & 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 0 & 0 & -0.5 & -0.5 & 0 \\ 0 & 0 & 0 & 0 & 2 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & -1 & 4 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 2 & 0 & 1 & 2.5 & 1.5 & -1 \end{bmatrix} \tag{3.31}$$

The resulting reduced 4×6 matrix is only a third of the size of the original matrix, but is still of full rank, and still has two degrees of freedom. However, while all significant information has been retained, the condition has decreased by half from 7.6 to 4.0 making the reduced matrix more robust to error fluctuations on the rates. That these two matrices are functionally equivalent, can be confirmed by comparing the results for the calculated fluxes of the oxidative phosphorylation and the ATP sink, which change only by 1.2 and 0.8 % respectively.

$$G^T = \begin{bmatrix} -2 & -1 & -\frac{1}{3} & 0 & 0 & 0 \\ -2 & -1 & \frac{4}{3} & 0 & -1 & 0 \\ -2 & 0 & \frac{2.5}{3} & 0 & 2.5 & -1 \\ 0 & 0 & -\frac{1}{6} & -1 & -0.5 & 0 \end{bmatrix} \tag{3.32}$$

Matrix is not of full rank: Let us consider a stoichiometric matrix in which the cofactor ATP is incorporated. If we would modify the reactions to also consider ADP, we would add another mass balance that is identical to the ATP mass balance multiplied by -1. Naturally, this matrix would not be of full rank anymore and the condition would increase tremendously. The condition of the matrix that is described later in the text in equation 3.39 would increase from 4.0 to 8.2×10^{16}. By iteratively removing one mass balance from the matrix and checking if the rank has not changed, we can easily determine the set of linearly dependent mass balances even for more complicated cases. This method can of course also be applied to the columns/fluxes of the matrix. Here it should be considered that many linearly dependent reactions often do not operate at the same time, therefore knowledge of the regulatory mechanisms of these pathways should be used to determine which reactions should be omitted (Stephanopoulos, Aristidou, and Nielsen 1998). An alternative approach could be to eliminate dependencies by considering the compartmentation of the reactions.

Cyclic pathways: Substrate cycles were initially thought to be a useless waste of ATP and therefore originally dubbed *futile cycles*. However, since their discovery, this view has changed, and substrate cycles have been recognized as important regularly mechanisms in metabolism that play a prominent role in glycolysis/gluconeogenesis. Their purpose has been described to lower the flux into a pathway as an "energetic price" to pay for the ability to rapidly switch between pathways (D. Voet, J. G.

Voet, and Pratt 2016). The substrate cycle of fructose-6-phosphate and fructose-1,6-bisphosphate can be described with the following equations

$$\text{glc} + \text{ATP} \rightarrow \text{fru-6-p} \tag{3.33}$$

$$\text{fru-6-p} + \text{ATP} \rightarrow \text{fru-1,6-p} \tag{3.34}$$

$$\text{fru-1,6-p} \rightarrow \text{fru-6-p} \tag{3.35}$$

$$\text{fru-1,6-p} \rightarrow \text{DHAP} + \text{G3P} \tag{3.36}$$

$$\rightarrow \text{ATP} \tag{3.37}$$

which lead to the following matrix form (excluding the substrate and two products)

$$G^T = \begin{bmatrix} -1 & -1 & 0 & 0 & 1 \\ 1 & -1 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 \end{bmatrix} \tag{3.38}$$

While such a matrix would seem reasonable at first inspection by having full rank and a condition of 4 problems may still arise. If we would want to solve such a system by only measuring v_1 and v_4 we would have linear dependencies. So even though the system might seem determined we would need to measure another system or remove one of the two reactions v_2 and v_3 based on prior knowledge of regulatory mechanisms. Luckily, such futile cycles are easily recognized and dealt with as long as the rank of G_c^T is checked to be full. If, however, this is not done, and the pseudo-inverse of G_c^T is calculated using the Moore-Penrose method, one might not even realise that the results calculated are not trustworthy! This would be the case, if one would use the unaltered MatLab code provided by Goudar et al. (2014).

Correlating mass balances: Let us consider again the matrix in equation 3.38 from the discussion of futile cycles. Let us assume that one would try to monitor this system in a more detailed fashion, by also including ADP together with ATP, so that one would formulate this new matrix G^T as

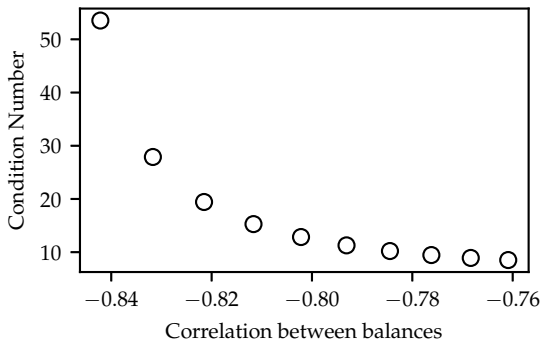

Figure 3.8: The condition number increases exponentially with an increased negative correlation between the mass balances of ATP and ADP. The stoichiometric factor X in equation 3.39 has been varied between -1.1 and -2.

$$G^T = \begin{bmatrix} -1 & -1 & 0 & 0 & 1 \\ 1 & -1 & 1 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 1 & 1 & 0 & 0 & X \end{bmatrix} \tag{3.39}$$

If one were to put X as -1, then the two balances of ATP and ADP would be linear dependent, and that would be noticeable by the rank and a high condition number. However, if one were to argue that the reaction from ADP to ATP would be not completely efficient and that some ADP would be "lost" due to side reactions, one could set X at any value lower than -1 to signify this. Therefore, the matrix would again be of full rank, even though the condition number would increase exponentially with the correlation factor between the two balances, as can be seen in figure 3.8.

It is therefore advisable to be very careful when including both forms of a cofactor, e.g., ADP and ATP, NAD^+ and NADH, Acetyl-CoA and CoA etc., and one should try, if possible, to formulate any losses in efficiency not through stoichiometric factors but through additional reactions in the intracellular network.

Insights from stoichiometric matrices

After now having discussed the properties and prerequisites of our matrix G^T, a few paragraphs will be written for additional information that can be extracted from the matrix itself or the intracellular fluxes calculated.

Calculating maximum theoretical yields: Calculating the theoretical yield can be as simple as solving for the stoichiometric coefficients in equation 3.40 by creating balances for C, N, H and degree of reduction (DoR). The maximum theoretical yield would then be defined as $Y = \frac{1}{a}$

$$\begin{aligned} a \cdot \text{Substrate} + b \cdot O_2 + c \cdot NH_3 \rightarrow \\ \text{Product} + d \cdot CO_2 + e \cdot H_2O \end{aligned} \tag{3.40}$$

This approach however ignores the demand of cofactors such as ATP, NADPH and NAD in biosynthesis, which would further decrease the theoretical yield closer to the true maximum yield. Even though this method yields better estimates, the correct formulation of the cofactor balances can easily become a complex task that is more error prone and therefore, a more systematic approach is preferable. MFA provides a suitable framework for such an approach, as one simply sets the flux of substrate to 100 mol/h and the fluxes of all other carbon delivering substrates and products to 0. The yield is then the fraction of product and substrate flux

$$Y = \frac{v_{Product}}{v_{Substrate}} \tag{3.41}$$

It should be noted that it can be still useful to calculate the theoretical yield through elemental balances first, as to have an upper boundary to compare the results of MFA for plausibility.

Metabolic rigidity: Yield itself is limited not by enzyme activity but by how much flux is channelled away from the intermediary branch points in the production pathway (Stephanopoulos and Vallino 1991). By varying the yield under similar conditions under which a theoretical maximum yield was observed, one can determine the so-called principle nodes. Principal nodes are defined as the nodes for which split ratios change with varying yields, i.e., the nodes that influence the yield by diverting fluxes away from product formation. After having defined these principle nodes, one can determine their rigidity through experimental flux perturbations. Flexible nodes allow for changes in split ratio from 0 to 1 and therefore usually do not need to be modified, while rigid nodes will be insusceptible to flux perturbations and would need to be either modified, or other fluxes will have to be diverted through other pathways to increase yields.

Elementary flux modes: If the columns of the stoichiometric matrix are linearly dependent, which will always be the case if the number of reactions exceeds the amount of metabolites, then a matrix R exists so that

$$G^T R = 0. \tag{3.42}$$

This matrix R is said to span the null space, so mathematically speaking, it is a basis of the solution space of the steady state assumptions. Therefore, each individual column of R and all linear combinations of the columns of R satisfy the steady state assumption $G^T v = 0$, so all v that fulfill the steady state condition can be described as a linear combination of the columns of R. Two difficulties to overcome for the use of this mathematical application were that such an R is not unique, i.e., there are a multitude of vector bases for G^T, and that not all of them allow for biological interpretation. As a consequence, and based on prior ideas put forward by Leiser and Blum (1987) and Fell (1993), Schuster and Hilgetag have defined the elementary modes of a stoichiometric matrix as the unique basis that fulfils the three conditions of steady state, sign restrictions and simplicity (S. Schuster and Hilgetag 1994; S. Schuster, Hilgetag, and R. Schuster 1996). The elementary modes can therefore be understood as functional units of the network, e.g., the sum of reactions that produce a certain product. This moves the analysis of stoichiometric networks away from single reactions to whole pathways. The elementary modes can be calculated by starting with the tableau $T^{(0)}$

$$T^{(0)} = \begin{bmatrix} I & 0 & G_{rev} \\ 0 & I & G_{irr} \end{bmatrix} = \begin{bmatrix} B^{(0)} \\ F^{(0)} \end{bmatrix} = \begin{bmatrix} I & G \end{bmatrix} \tag{3.43}$$

In an iterative fashion, for each column of G, we remove all stoichiometric factors of that column by linear combination of the rows of $T^{(0)}$. Linear combinations of the as reversible declared reactions are put into $B^{(x)}$, while linear combinations containing irreversible reactions are put into $F^{(x)}$. After this has been done n times for the n rows of G^T, we arrive at our solution in the final tableau $T^{(n)}$.

$$T^{(n)} = \begin{bmatrix} B^{(n)} \\ F^{(n)} \end{bmatrix} = \begin{bmatrix} b^{(k)} & 0 \\ f^{(k)} & 0 \end{bmatrix} \tag{3.44}$$

Where $b^{(k)}$ are the reversible flux modes and $f^{(k)}$ the irreversible. All possible elementary modes are therefore the set of $f^{(k)}$, $b^{(k)}$ and $-b^{(k)}$. It should be noted that this approach is conceptually quite similar to the pathway synthesis algorithm described by M. L. Mavrovouniotis and Stephanopoulos (1990) and M. Mavrovouniotis and Stephanopoulos (1992).

By finding the set of functional pathways from the stoichiometric matrix, one can apply this information to a wide array of problems: Provost and Bastin (2004) used this to create macro-reactions from the network and by assuming Monod-kinetics developed a possible framework for dynamic MFA. It can also be used for calculability analysis to determine which fluxes of a network need to be measured to be able to calculate a specific flux, even if the system at large might still be underdetermined (S. Schuster, Klamt, et al. 2002; Klamt, S. Schuster, and Gilles 2002). Additionally, just with the structural information alone one is able to deduce certain aspects of the regulation of the network such as relative transcription ratios (S. Schuster, Klamt, et al. 2002; Stelling et al. 2002). Furthermore, elementary flux modes can also be seen as an advancement of the calculation of theoretical yields (S. Schuster, Klamt, et al. 2002).

As a final paragraph to elementary flux modes, it should be noted that this is not the only approach to pathway analysis, the second in widespread use being extreme pathway analysis. Both use very similar concepts indeed, but extreme pathway analysis exchanges linear algebra and its equalities with the inequalities of convex analysis and replaces the basis of the flux space with the convex basis, the extreme rays that span a solution cone (Schilling, Letscher, and Palsson 2000). Extreme pathways are therefore a subset of elementary modes and are sometimes preferred to elementary flux modes, as their calculation can be computationally quite demanding (Klamt and Stelling 2002; Klamt and Stelling 2003). However, Klamt and Stelling believe that elementary flux modes are more useful and cite as a few reasons that extreme pathway analysis will lead to missing independent pathways and that removing or changing a single reaction will lead to recalculating of all extreme paths (2003).

Metabolite pools: In biological systems there exist groups of metabolites for which the sum of concentrations within this group is conserved (Park Jr. 1986; Park Jr. 1988; S. Schuster and Heinrich 1996):

$$\sum_{i=0}^{n} Met_i = const. \tag{3.45}$$

These groups containing a conserved moiety, i.e., a chemical entity that is conserved within, even for an open system, could be cofactor balances, such as the metabolites ATP, ADP and AMP. Another way to describe these dependencies between the material balances of the stoichiometric matrix would be

$$L \times G^T = 0 \tag{3.46}$$

which is equivalent to

$$G \times L = 0 \tag{3.47}$$

and therefore similar to equation 3.42 - $G^T \times R = 0$ - and the objectives of pathway analysis, where R represented the (right) null space, so the matrix L is called the left null space (Famili and Palsson 2003), or, more intelligibly, the conservation matrix (S. Schuster and Heinrich 1996). While R represents the possible linear combinations of fluxes under pseudo-steady state, L represents the linear combinations of concentrations that are allowed under these conditions. While the columns of R represent pathways, the columns of L represent "metabolic pools" (Famili and Palsson 2003). A difference between R and L is that while R might contain negative factors if some reactions are considered reversible,

the factors in L must all be non-negative (S. Schuster and Heinrich 1996). Just as for R, there exist a variety of methods for calculating the conservation matrix L, for instance by the use of convex analysis (S. Schuster and Heinrich 1996; Famili and Palsson 2003), linear algebra (Reder 1988; Cornish-Bowden and Hofmeyr 2002) or other approaches (Nikolaev, Burgard, and Maranas 2005). Metabolic pools understood as conservation relationships are said to provide a meaningful basis for understanding cell-signalling (Famili and Palsson 2003) and metabolic regulation by assuming metabolic pools of enzymes and its effectors (Jamshidi and Palsson 2010).

Sensitivity analysis

In its essence, sensitivity analysis deals with the effect a change in that model has on model outputs. A "change in model" is here defined very broadly and can relate to structure, parameters or inputs alike. While it is often used to the study of how uncertainty in model parameters propagate to model outputs (Sin, Gernaey, and Lantz 2009), these techniques can also be applied to many other problems such as model reduction, detection of over- and under-fitting, alternative model formulation (Saltelli et al. 2004), identifiability analsis (Miao et al. 2011) and covariance structure analysis (Tanaka and Watadani 1992). In general terms, one can discern between local and global sensitivity analysis. Local methods only determine the sensitivity at one point of the parameter's space and each parameter is investigated on its own, while global methods investigate the whole multi-dimensional parameter space (Saltelli et al. 2004). Even for simple linear regression, the field of sensitivity analysis spans enough topics to fill a whole book (Chatterjee and Hadi 2009), however, in this section only the most fundamental terms will be explained. The local sensitivity of input θ on output y, is simply defined as the derivative

$$S_{y,\theta} = \frac{\partial y}{\partial \theta} \tag{3.48}$$

Some take issue with the often confusing units and values of the sensitivity (Gujer 2008) and use either the absolute-relative sensitivity (absolute change of y per relative change of θ)

$$S^{a,r}_{y,\theta} = \theta \frac{\partial y}{\partial \theta} \tag{3.49}$$

or the relative sensitivity

$$S^{r}_{y,\theta} = \frac{\theta \partial y}{y \partial \theta} = \frac{\partial \ln y}{\partial \ln \theta} \tag{3.50}$$

which is sometimes also called the logarithmic sensitivity (Gheshlaghi et al. 2007). The logarithmic sensitivity of metabolic fluxes are also mathematically identical to the flux control coefficients of metabolic control analysis (Bailey 1998; Stephanopoulos, Aristidou, and Nielsen 1998).

For the specific case of MFA with underdetermined systems, the concept of *shadow prices* can be applied. The shadow price is the sensitivity of the optimization function to a change in measured fluxes with the optimization function usually being a flux that is thought to be maximised by the organism during growth, so usually the biomass production or product formation (Stephanopoulos, Aristidou, and Nielsen 1998; Gheshlaghi et al. 2007).

References

Publications of special significance and value have their titles highlighted in boldface and have been complemented by an explanatory comment.

Allaire, G. (2007). *Numerical analysis and optimization: an introduction to mathematical modelling and numerical simulation*. Oxford university press.

Bader, F. G. (1978). "Analysis of double-substrate limited growth". In: *Biotechnology and bioengineering* 20.2, pp. 183–202. DOI: 10.1002/bit.260200203.

Bailey, J. E. (1998). "**Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities**". In: *Biotechnology progress* 14.1, pp. 8–20. DOI: 10.1021/bp9701269.
This paper gives a good overview over the history of modelling in bioengineering and even after more than 20 years provides useful input for the prospective modeller. Additionally, this publications deals with a few all to common misconceptions such as "shear damage" and what really makes growth factors more effective.

Barata, J. C. A. and M. S. Hussein (2012). "The Moore–Penrose pseudoinverse: A tutorial review of the theory". In: *Brazilian Journal of Physics* 42.1-2, pp. 146–165. DOI: 10.1007/s13538-011-0052-z.

Basheer, I. A. and M. Hajmeer (2000). "**Artificial neural networks: fundamentals, computing, design, and application**". In: *Journal of microbiological methods* 43.1, pp. 3–31. DOI: 10.1016/S0167-7012(00)00201-3.
While this publication might be considered outdated in such a fast growing field as A.I., it still gives a very detailed overview of strategies and considerations when dealing with artificial neural networks. The principles are applied to modelling microbial growth on different conditions such as temperature, pH and salt concentrations.

Bellu, G. et al. (2007). "DAISY: A new software tool to test global identifiability of biological and physiological systems". In: *Computer methods and programs in biomedicine* 88.1, pp. 52–61. DOI: 10.1016/j.cmpb.2007.07.002.

Best, JB (1955). "The inference of intracellular properties from observed kinetic data". In: *J. Cell Comp. Physiol* 46, pp. 1–27. DOI: 10.1002/jcp.1030460102.

Blackman, F. F. (1905). "Optima and limiting factors". In: *Annals of botany* 19.74, pp. 281–295. DOI: 10.1093/oxfordjournals.aob.a089000.

Chatterjee, S. and A. S. Hadi (2009). *Sensitivity analysis in linear regression*. Vol. 327. John Wiley & Sons. ISBN: 978-0-470-31742-6.

Chen, Y. and J. Feng (2014). "Efficient method for Moore-Penrose inverse problems involving symmetric structures based on group theory". In: *Journal of Computing in Civil Engineering* 28.2, pp. 182–190. DOI: 10.1061/(ASCE)CP.1943-5487.0000266.

Contois, D. E. (1959). "Kinetics of bacterial growth: relationship between population density and specific growth rate of continuous cultures". In: *Microbiology* 21.1, pp. 40–50. DOI: 10.1099/00221287-21-1-40.

Cornish-Bowden, A. and J.-H. S. Hofmeyr (2002). "The role of stoichiometric analysis in studies of metabolism: an example". In: *Journal of Theoretical Biology* 216.2, pp. 179–191. DOI: 10.1006/jtbi.2002.2547.

Dabes, J. N., R. K. Finn, and C. R. Wilke (1973). "Equations of substrate-limited growth: The case for blackman kinetics". In: *Biotechnology and Bioengineering* 15.6, pp. 1159–1177. DOI: 10.1002/bit.260150613.

Dick, E. J. (2004). "Beyond 'lognormal versus gamma': discrimination among error distributions for generalized linear models". In: *Fisheries Research* 70.2-3, pp. 351–366. DOI: 10.1016/j.fishres.2004.08.013.

Famili, I. and B. Ø. Palsson (2003). "**The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools**". In: *Biophysical journal* 85.1, pp. 16–26. DOI: 10.1016/S0006-3495(03)74450-6.
The principles of extreme pathway analysis are not applied to determine pathways from reactions, but conserved metabolite pools from the metabolites.

Fell, D. A. (1993). "The analysis of flux in substrate cycles". In: *Modern Trends in Biothermokinetics*. Springer, pp. 97–101. DOI: 10.1007/978-1-4615-2962-0_17.

Ferenci, T. (1999). "Growth of bacterial cultures' 50 years on: towards an uncertainty principle instead of constants in bacterial growth kinetics". In: *Research in microbiology* 150.7, pp. 431–438. DOI: 10.1016/S0923-2508(99)00114-X.

Gheshlaghi, R. E. Z. A. et al. (2007). "Metabolic flux analysis for optimizing the specific growth rate of recombinant *Aspergillus niger*". In: *Bioprocess and biosystems engineering* 30.6, pp. 397–418. DOI: 10.1007/s00449-007-0136-x.

Goudar, C. T. et al. (2014). "**Metabolic flux estimation in mammalian cell cultures**". In: *Animal Cell Biotechnology*. Springer, pp. 193–209. DOI: 10.1007/978-1-62703-733-4_13.
Short book chapter that explains the principles of MFA in an easy to follow step-by-step approach. Especially for the calculation of the consistency index it can be a valuable source of data for unit testing of implemented code. However, it is quite light on underlying theory and should not be seen as an alternative to the more comprehensive volume by Stephanopoulos, Aristidou & Nielsen (1998) but only as a supplement or brush-up of the basics.

Gujer, W. (2008). "Parameter Identification, Sensitivity and Error Propagation". In: *Systems Analysis for Water Technology*, pp. 257–313. DOI: 10.1007/978-3-540-77278-1_12.

Heijnen, J. J. and B. Romein (1995). "Derivation of kinetic equations for growth on single substrates based on general properties of a simple metabolic network". In: *Biotechnology progress* 11.6, pp. 712–716. DOI: 10.1021/bp00036a018.

Jamshidi, N. and B. Ø. Palsson (2010). "**Mass action stoichiometric simulation models: incorporating kinetics and regulation into stoichiometric models**". In: *Biophysical journal* 98.2, pp. 175–185. DOI: 10.1016/j.bpj.2009.09.064.
A novel approach that tries to combine the data generated from the different -omics to provide a framework for comprehensive modelling of cells.

Jørgensen, H. S. et al. (1995). "Metabolic flux distributions in *Penicillium chrysogenum* during fed-batch cultivations". In: *Biotechnology and Bioengineering* 46.2, pp. 117–131. DOI: 10.1002/bit.260460205.

Klamt, S., S. Schuster, and E. D. Gilles (2002). "Calculability analysis in underdetermined metabolic networks illustrated by a model of the central metabolism in purple nonsulfur bacteria". In: *Biotechnology and Bioengineering* 77.7, pp. 734–751. DOI: 10.1002/bit.10153.

Klamt, S. and J. Stelling (2002). "Combinatorial complexity of pathway analysis in metabolic networks". In: *Molecular biology reports* 29.1-2, pp. 233–236. DOI: 10.1023/A:1020390132244.

– (2003). "**Two approaches for metabolic pathway analysis?**" In: *Trends in biotechnology* 21.2, pp. 64–69. DOI: 10.1016/S0167-7799(02)00034-3.
This publication gives a good overview over differences, similarities and strenghts of elementary flux modes and extreme pathway analysis.

Koch, A. L. (1998). "The Monod model and its alternatives". In: *Mathematical modeling in microbial ecology*. Springer, pp. 62–93. DOI: 10.1007/978-1-4615-4078-6_4.

Kovárová-Kovar, K. and T. Egli (1998). "Growth kinetics of suspended microbial cells: from single-substrate-controlled growth to mixed-substrate kinetics". In: *Microbiol. Mol. Biol. Rev.* 62.3, pp. 646–666. DOI: 10.1128/MMBR.62.3.646-666.1998.

Kroll, P. et al. (2017). "Workflow to set up substantial target-oriented mechanistic process models in bioprocess engineering". In: *Process Biochemistry* 62, pp. 24–36. DOI: 10.1016/j.procbio.2017.07.017.

L'Ecuyer, P. and R. Simard (2007). "TestU01: AC library for empirical testing of random number generators". In: *ACM Transactions on Mathematical Software (TOMS)* 33.4, pp. 1–40. DOI: 10.1145/1268776.1268777.

Leiser, J. and J. J. Blum (1987). "On the analysis of substrate cycles in large metabolic systems". In: *Cell biophysics* 11.1, pp. 123–138. DOI: 10.1007/BF02797119.

Mavrovouniotis, M. and G. Stephanopoulos (1992). "Synthesis of biochemical production routes". In: *Computers & chemical engineering* 16.6, pp. 605–619. DOI: 10.1016/0098-1354(92)80071-G.

Mavrovouniotis, M. L. and G. Stephanopoulos (1990). "Computer-aided synthesis of biochemical pathways". In: *Biotechnology and bioengineering* 36.11, pp. 1119–1132. DOI: 10.1002/bit.260361107.

Meeker, W. Q. and L. A. Escobar (1995). "Teaching about approximate confidence regions based on maximum likelihood estimation". In: *The American Statistician* 49.1, pp. 48–53. DOI: 10.1080/00031305.1995.10476112.

Miao, H. et al. (2011). "On identifiability of nonlinear ODE models and applications in viral dynamics". In: *SIAM review* 53.1, pp. 3–39. DOI: 10.1137/090757009.

Monod, J. (1941). "Recherches sur la croissance des cultures bacteriennes". In:

Nikolaev, E. V., A. P. Burgard, and C. D. Maranas (2005). "Elucidation and structural analysis of conserved pools for genome-scale metabolic reconstructions". In: *Biophysical journal* 88.1, pp. 37–49. DOI: 10.1529/biophysj.104.043489.

Nyberg, G. B. et al. (1999). "Metabolism of peptide amino acids by Chinese hamster ovary cells grown in a complex medium". In: *Biotechnology and bioengineering* 62.3, pp. 324–335. DOI: 10.1002/(SICI)1097-0290(19990205)62:3%3C324::AID-BIT9%3E3.0.CO;2-C.

O'Neill, M. E. (2014). "PCG: A family of simple fast space-efficient statistically good algorithms for random number generation". In: *ACM Transactions on Mathematical Software*. DOI: 10.1145/0000000.000000.

Papoulis, A. and S. U. Pillai (2002). *Probability, random variables, and stochastic processes*. Tata McGraw-Hill Education.

Park Jr., D. J. M. (1986). "The complete stoichiometer". In: *Computer methods and programs in biomedicine* 22.3, pp. 293–301. DOI: 10.1016/0169-2607(86)90006-4.

– (1988). "Positive compositional algorithms in chemical reaction systems". In: *Computers & chemistry* 12.2, pp. 175–188. DOI: 10.1016/0097-8485(88)85015-0.

Powell, E. O. (1967). "The growth rate of microorganisms as a function of substrate concentration". In: *Microbial physiology and continuous culture* 3.

Press, W. H. et al. (2007). *Numerical recipes 3rd edition: The art of scientific computing*. Cambridge university press. ISBN: 9780521880688.

Provost, A. and G. Bastin (2004). "**Dynamic metabolic modelling under the balanced growth condition**". In: *Journal of Process Control* 14.7, pp. 717–728. DOI: 10.1016/j.jprocont.2003.12.004.
A detailed description of the general dynamic metabolic model, metabolic flux analysis, and how to generate macro-reactions from the stoichiometric model. Uses simple Monod kinetics for dynamic modelling of the macro-reactions, which were dereived from pathway analysis by using elementary flux modes. Based on the assumptions of balanced and exponential growth.

Raue, A., V. Becker, et al. (2010). "Identifiability and observability analysis for experimental design in nonlinear dynamical models". In: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 20.4, p. 045105. DOI: 10.1063/1.3528102.

Raue, A., J. Karlsson, et al. (2014). "Comparison of approaches for parameter identifiability analysis of biological systems". In: *Bioinformatics* 30.10, pp. 1440–1448. DOI: 10.1093/bioinformatics/btu006.

Raue, A., C. Kreutz, et al. (2009). "**Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood**". In: *Bioinformatics* 25.15, pp. 1923–1929. DOI: 10.1093/bioinformatics/btp358.
Explanation on how to use the profile likelihood for parameter identifiability with examples.

Reder, C. (1988). "Metabolic control theory: a structural approach". In: *Journal of theoretical biology* 135.2, pp. 175–201. DOI: 10.1016/S0022-5193(88)80073-0.

Rickenberg, H. V. et al. (1956). "La galactoside-permease d'*Escherichia coli*". In: *Ann. Inst. Pasteur* 91, pp. 829–857.

Robinson, D. and C. Atcitty (1999). "Comparison of quasi-and pseudo-Monte Carlo sampling for reliability and uncertainty analysis". In: *40th Structures, Structural Dynamics, and Materials Conference and Exhibit*, p. 1589. DOI: https://doi.org/10.2514/6.1999-1589.

Saltelli, A. et al. (2004). *Sensitivity analysis in practice: a guide to assessing scientific models*. Vol. 1. Wiley Online Library.

Schilling, C. H., D. Letscher, and B. Ø. Palsson (2000). "*Theory for the systemic definition of metabolic pathways and their use in interpreting metabolic function from a pathway-oriented perspective*". In: *Journal of theoretical biology* 203.3, pp. 229–248. DOI: 10.1006/jtbi.2000.1073.
An alternative to elementary flux modes to approach pathway analysis, extreme pathway analysis uses convex analysis as its mathematical basis.

Schuster, S. and R. Heinrich (1996). *The regulation of cellular systems*. DOI: 10.1007/978-1-4613-1161-4.

Schuster, S. and C. Hilgetag (1994). "**On elementary flux modes in biochemical reaction systems at steady state**". In: *Journal of Biological Systems* 2.02, pp. 165–182. DOI: 10.1142/S0218339094000131.
In this publication a simple definition of elementary flux modes is given, which gives rise to an unambigous definition of them. This publication marks the step from metabolic flux analysis to pathway analysis.

Schuster, S., C. Hilgetag, and R. Schuster (1996). "Determining elementary modes of functioning in biochemical reaction networks at steady state". In: *Biomedical and Life Physics*. Springer, pp. 101–114. DOI: 10.1007/978-3-322-85017-1_9.

Schuster, S., S. Klamt, et al. (2002). "Use of network analysis of metabolic systems in bioengineering". In: *Bioprocess and Biosystems Engineering* 24.6, pp. 363–372. DOI: 10.1007/s004490100253.

Sedoglavic, A. (2002). "A probabilistic algorithm to test local algebraic observability in polynomial time". In: *Journal of Symbolic Computation* 33.5, pp. 735–755. DOI: 10.1006/jsco.2002.0532.

Sin, G., K. V. Gernaey, and A. E. Lantz (2009). "Good modeling practice for PAT applications: propagation of input uncertainty and sensitivity analysis". In: *Biotechnology progress* 25.4, pp. 1043–1053. DOI: 10.1002/btpr.166.

Singhee, A. and R. A. Rutenbar (2010). "Why quasi-monte carlo is better than monte carlo or latin hypercube sampling for statistical circuit analysis". In: *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems* 29.11, pp. 1763–1776. DOI: 10.1109/TCAD.2010.2062750.

Stelling, J. et al. (2002). "Metabolic network structure determines key aspects of functionality and regulation". In: *Nature* 420.6912, pp. 190–193. DOI: 10.1038/nature01166.

Stephanopoulos, G., A. A. Aristidou, and J. Nielsen (1998). ***Metabolic Engineering: principles and methodologies***. ISBN: 9780080536286.
The value of this book cannot be overstated. Chapters 3 and 8 of this educational book give a very good overview of the theory and applications behind MFA. This volume should be considered the go-to work if one wishes to accustume themselves to MFA, MCA and metabolic engineering in general.

Stephanopoulos, G. and J. J. Vallino (1991). "Network rigidity and metabolic engineering in metabolite overproduction". In: *Science* 252.5013, pp. 1675–1681. DOI: 10.1126/science.1904627.

Strang, G. (2005). "Linear Algebra and its Applications". In: *Cole, Belmont, CA, USA*.

Tanaka, Y. and S. Watadani (1992). "Sensitivity analysis in covariance structure analysis with equality constraints". In: *Communications in statistics-theory and methods* 21.6, pp. 1501–1515. DOI: 10.1080/03610929208830861.

Tellinghuisen, J.l (2001). "Statistical error propagation". In: *The Journal of Physical Chemistry A* 105.15, pp. 3917–3921. DOI: 10.1021/jp003484u.

Van der Heijden, R. T. J. M. et al. (1994). "Linear constraint relations in biochemical reaction systems: I. Classification of the calculability and the balanceability of conversion rates". In: *Biotechnology and Bioengineering* 43.1, pp. 3–10. DOI: 10.1002/bit.260430103.

Voet, D., J. G. Voet, and C. W. Pratt (2016). *Fundamentals of biochemistry: life at the molecular level*. John Wiley & Sons. ISBN: 978-1-119-42357-7.

Walter, E. (2014). *Identifiability of parametric models*. Elsevier. ISBN: 978-1-483-15595-1.

Wang, N. S. and G. Stephanopoulos (1983). "Application of macroscopic balances to the identification of gross measurement errors". In: *Biotechnology and bioengineering* 25.9, pp. 2177–2208. DOI: 10.1002/bit.260250906.

Wang, X. (2007). "Effect of small rank modification on the condition number of a matrix". In: *Computers & Mathematics with Applications* 54.6, pp. 819–825. DOI: 10.1016/j.camwa.2006.05.031.

Part II

Results

Chapter 4

Achievements

Summary

To answer the scientific questions stated previously, we had to determine at first which bound amino acids are released and which kinetics can describe their release. In fungi a common strategy to metabolise insoluble substrates is to release extracellular enzymes which degrade the substrates into soluble compounds. Cultures of *Penicillium chrysogenum* were cultivated in a bioreactor. From these experiments, the supernatant at beginning and end of the Fed-Batch was incubated with fresh corn steep liquor and the release of free amino acids was monitored via high pressure liquid chromatography (HPLC). From these curves, the simplest identifiable model was determined which sufficiently described amino acid release.

Using these kinetics, historic fed-batch fermentation data was reevaluated by recalculating the uptake and production rates of amino acids. These rates were used as input for a metabolic flux model developed from published models and the Kyoto encyclopaedia of genes and genomes. From a local sensitivity analysis, new hypothesis for limitating substrates could be generated. These results were confirmed with a new fermentation run where methionine was pulsed after its depletion.

Findings and Impact

BASED on the current state of the art and the scientific questions formulated in the previous chapter, it was possible to establish the following knowledge in this thesis:

What are the release kinetics of amino acids from corn steep liquor?

It could be established that release kinetics are not negligible and that its neglect can lead to wrong estimation of uptake rates. Furthermore, it was discovered that the release kinetics of amino acids vary. In general terms amino acids could be categorized in those that i) show no release at all, ii) are completely released during batch-phase, iii) are partly released during batch-phase and iv) are only released in the fed-batch phase. Conventional wisdom that corn steep liquor only affects the batch-phase could be disproven.

What is the role of amino acids on penicillin formation? Which are limiting over process time and why?

As penicillin is a peptide created from the amino acids α-aminoadipic acid, valine and cysteine, it is only natural to suspect them to affect its production. It could be shown that α-aminoadipic acid and valine both have only a marginal effect, while penicillin shows the greatest sensitivity to the sulphur carrying amino acids cysteine and methionine. This is due to the high NADPH demand of penicillin production, which already in previous studies have been shown to be the factor limiting the yield

(van Gulik et al., 2001). As sulphur utilization also consumes NADPH, the uptake of sulphur containing amino acids is able to circumvent this energetic price and is therefore able to spend the NADPH in other reaction steps.

What is the role of amino acids on biomass formation? Which are limiting over process time and why?

For biomass production the difference in effects are less pronounced with tryptophan, alanine, glutamine and asparagine being the amino acids with the greatest sensitivity. The sensitivities change marginally over time, suggesting that no new limitations arise during fed-batch fermentation.

How do energetic demands guide and limit penicillin formation?

As already mentioned before, the role of amino acid and energetic limitations are both deeply intertwined in the biosynthesis of penicillin. NADPH, and not ATP or specific carbon species, is the limiting factor for the process as NADPH is used in many steps including sulphur and nitrogen utilization and for the synthesis of different amino acids. This leads to the conclusion that the use of alternative sulphur and nitrogen sources, which are energetically not as demanding as sulphate and ammonia, could lead to increased yields. Alternatively, genetically modifying *Penicillium chrysogenum* by introducing alternative routes of sulphur and nitrogen utilization could have an identical effect.

Do metabolic shifts occur during fermentation? Can these shifts be used for process characterization?

There seem to be no fundamental shifts in metabolism during fed-batch fermentation, even though, after pulsing methionine the calculated intracellular fluxes estimate a decrease in fluxes of the pentose phosphate pathway. Otherwise no further changes or limitations could be detected.

This should act as a proof of concept that even for well-established processes metabolic flux analysis can lead to improved process knowledge and in turn to knowledge-based definition of critical process and material attributes. It can serve as a valuable method in the toolbox of the bioprocess engineer. However, conversely there are many things and details to consider that also make it a sometimes difficult to use tool. The formulation of an appropriate stoichiometric matrix can be a daunting task and its verification and validation even more so.

Chapter 5

Full Manuscripts

In-depth characterization of the raw material corn steep liquor and its bioavailability in bioprocesses of *Penicillium chrysogenum*

This manuscript was published in *Process Biochemistry*.

Citation: Hofer, A., **Hauer, S.**, Kroll, P., Fricke, J., & Herwig, C. (2018). In-depth characterization of the raw material corn steep liquor and its bioavailability in bioprocesses of Penicillium chrysogenum. Process biochemistry, 70, 20-28.

Submitted: 16th February 2018

Available online: 18th April 2018

DOI: 10.1016/j.procbio.2018.04.008

My contribution: Performed the experiments on enzymatic release together with Alexandra Hofer. Determined a kinetic model for enzyme release suitable for the data.

In-depth characterization of the raw material corn steep liquor and its bioavailability in bioprocesses of *Penicillium chrysogenum*

Alexandra Hofer[a], Stefan Hauer[a], Paul Kroll[a,b], Jens Fricke[a,b], Christoph Herwig[a,b]

[a] Research Area Biochemical Engineering, Institute of Chemical, Environmental and Bioscience Engineering, Vienna University of Technology, Gumpendorferstrasse 1a – 166/4, A-1060 Vienna, Austria
[b] CD Laboratory on Mechanistic and Physiological Methods for Improved Bioprocesses, Vienna University of Technology, Gumpendorferstrasse 1a – 166/4, A-1060 Vienna, Austria

Abstract

ALTHOUGH corn steep liquor (CSL) is a commonly applied raw material in biopharmaceutical production processes, its mechanistic impact on process performance is still not completely unraveled. It is a multicomponent and multiphase matrix, including free and bound ingredients. Hence, changes in bioavailability during cultivations and their effect on the process are challenging to determine.

In this study we want to give a holistic overview of CSL considering both phases as well as free and bound ingredients. With the example of amino acids bioavailability in CSL was assessed with respect to i) initial bioavailability ii) solubilization release in media and iii) enzymatic release by *Penicillium chrysogenum* during a fermentation process. Ten amino acids showed an enzymatic release including arginine, which is used for biomass formation, and valine, a precursor amino acid for penicillin production. Applicability of these results was demonstrated on a penicillin process for both these amino acids showing opposed meaning of specific rates depending on inclusion of release kinetics.

Overall, the detailed characterization of CSL with high analytical quality and the consideration of release kinetics resulted in a huge impact on the interpretation of the process with respect to fungal physiology and metabolism.

Introduction

CORN STEEP LIQUOR (CSL) is widely used as media supplement in diverse biopharmaceutical production processes (Kanzaki et al. 1967; Kona, Qureshi, and Pai 2001; Rivas et al. 2004; Silveira et al. 2001; Tornqvist and Peterson 1956). It is commonly added as cheap N-, C-, vitamin or trace element source as an alternative to expensive, defined media ingredients. The most prominent example of a CSL supplemented bioprocess is the production process of antibiotics by *Penicillium chrysogenum*. In this process, CSL was first added in the 1940s and resulted in an increase in productivity of about 20% (Bowden and Peterson 1946). Generally, the use of CSL holds many advantages: CSL is a byproduct of the wet-milling process (Hull et al. 1996). Hence, it is available at low-cost and as a waste product, that can be reused for the production of valuable products. Additionally, CSL originates from corn, hence, it is enriched with a wide variety of nutrients and is therefore able to increase growth and productivity when used in media of bioprocesses. On the other hand, the biological origin brings about some disadvantages as well. CSL is subject to a high and unpredictable lot-to-lot variability, which can originate from season, origin of the corn and the wet-milling process itself (Liggett and Koffler 1948). Furthermore, CSL is a multiphase system consisting of a liquid phase that is miscible with water and a solid phase which shows a low solubility in water. So far, this multiphase system has very scarcely been characterized leading to an ignorance of the actual bioavailability of the nutritious substances in CSL. In terms of the process analytical technologies (PAT) initiative, it is necessary to characterize critical raw material attributes (CMAs) and control the raw materials used in order to guarantee a safe product (FDA 2004). For complex raw materials such as CSL, this is a demanding task as it is

difficult to analyze the matrix and forecast the actual bioavailability of substances during the process.

So far, studies about CSL have mainly focused on the characterization of the liquid phase of the multiphase system, neglecting any bioavailable substances in the solid phase (Xiao et al. 2013). The impact of the complex nutrient source on process performance has just been investigated as black box system, i.e. no detailed characterization of the behavior of the various ingredients of CSL during the fermentation has been made. However, correlations between CSL quality and process performance attributes, i.e. product concentration or maximal biomass (Gao and Yuan 2011; Hofer, Kamravamanesh, et al. 2018), have been established. The FDA suggests to replace these black box systems by process knowledge in order to ensure product quality.

The focus of this study is to deliver a detailed characterization of CSL with respect to amino acids, organic acids, sugars, vitamins, trace elements and minerals. Furthermore, bioavailability of the ingredients is of fundamental importance in order to understand the impact on a bioprocess. For antibiotic production and filamentous growth amino acids are of utmost importance. Hence, bioavailability of the raw CSL as well as bioavailability through enzymatic release by *P. chrysogenum* was evaluated focusing on CSL as N-source, i.e. amino acid source. The underlying hypothesis of the availability of amino acids in CSL is shown in figure 5.1. Amino acids can be released by solubilization effects in the media or enzymatically by the fungi.

In this study we deliver a holistic overview of CSL with the focus on amino acids as nutrient source. Amino acids occur in initially bioavailable and bound form, hence, the change of bioavailability of the single amino acids due to enzymatic release over time is examined during processes of *P. chrysogenum*. We aim to show that the consideration of the enzymatic release kinetics has a huge impact on the interpretation of the process with respect to fungal physiology and metabolism.

Generally, CSL is widely used in biopharmaceutical production processes, thus, the generated knowledge can be applied for increasing the understanding of metabolic behavior in various processes.

Materials and methods

Chemicals and reagents

All chemicals and substances were of analytical grade and purchased from either Carl Roth (Karlsruhe, Germany) or Sigma Aldrich (St. Louis, MO, USA). For analytical methods ultra-pure water was used, derived from a Milli-Q system from Merck Millipore (Billerica, MA, USA).

Analytical methods

Vitamins were analyzed according to Hofer and Herwig (2017).

Amino acids, organic acids, sugars, sugar alcohols, trace elements and minerals were analyzed according to Hofer, Kamravamanesh, et al. (2018). In order to evaluate the substances that are bound in CSL, extraction procedures had to be performed for the different analysis, which were described before (Hofer and Herwig 2017; Hofer, Kamravamanesh, et al. 2018).

CSL samples

CSL samples originated from a European manufacturer, produced from dried corn. CSL samples are naturally inhomogeneous and consist of two immiscible phases (a solid and a liquid phase). For the extraction of the specific components (amino acids and vitamins) the samples were dried at 95 °C for 72 h. Otherwise the homogenized sample was used for analysis.

Kinetic experiments

The medium used for the release experiments contained CSL, $CaCO_3$ and Glucose.

Solubilization experiments were performed in 500 mL shake flasks. The medium containing CSL was shaken at 200 rpm at 25 °C for 96 h. Samples were taken every 24 h and analyzed by HPLC for amino acid concentration.

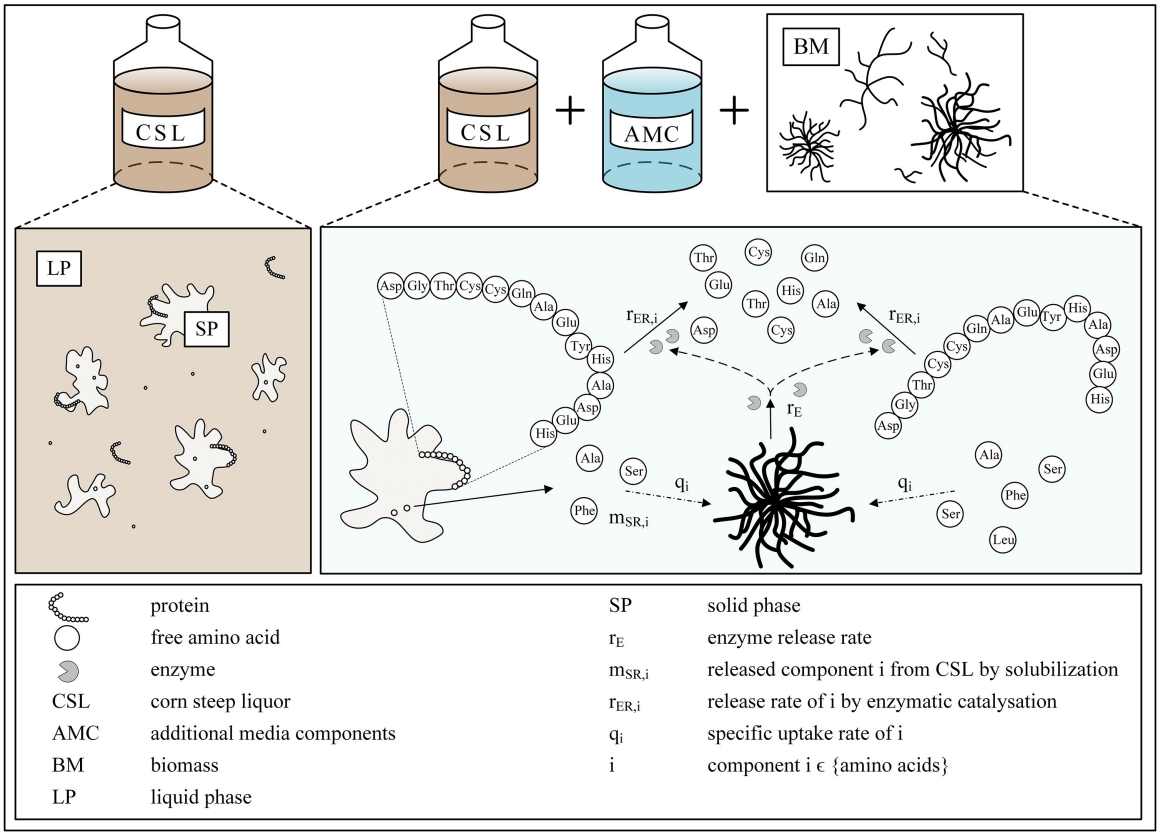

Figure 5.1: CSL can be understood as a multiphase system which consists of a liquid phase (LP) and a solid phase (SP). Both phases include free, thus bioavailable amino acids and protein-bound, thus, inaccessible amino acids (left plot). In a bioprocess these phases are mixed with additional media component (AMC) as well as with biomass (BM). Additionally to the soluble and bioavailable amino acids from LP and SP ($m_{SR,i}$), the fungus secretes proteases (r_i) which are able to degrade the CSL proteins ($r_{ER,i}$) and make the amino acids accessible as nutrients (q_i) (right plot).

For the enzymatic release experiments by Penicillium chrysogenum, a penicillin cultivation was performed as described in 2.6 Bioreactor cultivations. A sample of the batch end was centrifuged in a sterile manner 3 times at 4800 rpm and 4 °C for 10 min. 100 mL of supernatant were transferred by syringe in 150 mL sterile medium containing CSL (duplicate experiment) LP or SP in a 300 mL DASbox multibioreactor system (Eppendorf AG, Germany). Stirrer speed was controlled at 400 rpm and aeration was set to 1 vvm. The processes were run for 70 h and sampling was achieved by a Numera system, i.e. an automated sampling system (Securecell, Switzerland). The same procedure was performed with a fed-batch sample that was taken after 60 h of cultivation. For the batch supernatant processes samples were taken every hour for the first 20 h, followed by a sampling interval of 6 h. For the fed-batch supernatant processes, samples were taken every 6 h. The samples were cooled at 4 °C and analyzed by HPLC at the end of the processes.

Release kinetics evaluation

Under the assumption of an advanced Monod kinetic with only partially available substrate, the

measured data were smoothed. From this smoothed data the maximal release rate (r_{max}) and the half-life time ($t_{1/2} = t(r = r_{max}/2)$) were calculated.

Bioreactor cultivations

Cultivations were performed in a DASGIP multi-bioreactor system (Eppendorf AG, Germany) with a working volume of 2 L each. The initial batch and fed-batch volumes were 2 L and 1.5 L, respectively.

Penicillium chrysogenum was cultivated at controlled pH (6.5) – only during fed-batch –, dO_2 (>40%) and temperature (25 °C). Aeration was kept at 1 vvm. Dissolved oxygen and pH were measured using the optical and potentiometric probes, respectively (Hamilton, Switzerland). DO_2 was controlled via adjustment of stirrer speed and increase of the O_2 content in the inflow air and pH via 20% (w/v) KOH or 15% (v/v) H_2SO_4 addition. CO_2 and O_2 concentrations were analyzed with an off-gas analyzer (module DASGIP GA4) by a ZrO_2 sensor for O_2 and an infrared sensor for CO_2 analysis.

The batch was carried out on a complex bioreactor medium, including glucose and sucrose as carbon source, similar to the one described by Posch et al. (2013) without pH control. The end of the batch was indicated by an increase in pH by 0.5. Fed-batch was performed on a defined medium (similar to Posch et al. (2013)) which was inoculated by 10% of the batch broth. Three feeds were added to the fed-batch, namely glucose, ammonia – as nitrogen source – and a penicillin-precursor feed, via peristaltic pumps (module DASGIP M8, Eppendorf AG, Germany).

Cell dry weight was determined by centrifugation of 5 mL cultivation broth at 4800 rpm, 4 °C for 10 min. The pellet was resuspended once in deionized water and centrifuged again, followed by drying of the cell pellet at 95 °C for 72 h.

Results and discussion

In-depth characterization of CSL

CSL was characterized for 55 specific components (21 amino acids, 5 organic acid, 8 reducing sugars, 7 water-soluble vitamins and 14 trace elements/minerals). For a standard CSL quality sample over 80% of its composition could be specified by the established analytical methods (Table 5.1). The main ingredient is lactate which can be used as C-source in bioprocesses and amino acids, typical N-sources in more complex processes. The amount of the available amino acids is distributed in a range between 0.8 and 44.3 mg g^{-1}. The main amino acids are arginine, alanine and glutamic acid with 44.3, 35.7 and 42.0 mg g^{-1}.Vitamins, minerals and trace elements are available in very small amounts, mostly bound to proteins or other molecules. Other C-sources than lactate, namely sugars and organic acids, are present in negligible amounts for application in bioprocesses.

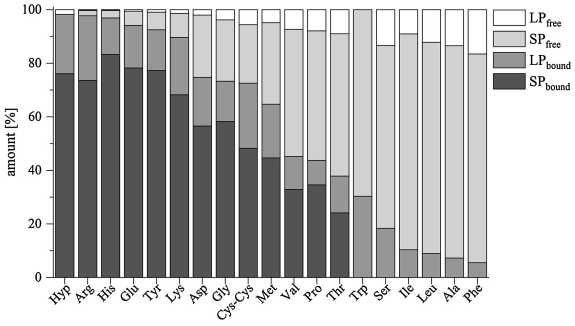

Figure 5.2: CSL was analyzed for the occurrence of free and bound amino acids in its liquid and solid phase, respectively. The amino acids are in order, starting with the one that shows the highest percentage of bound amino acid ending with those amino acids that are almost completely available in their free form.

In this study the focus lies on the application of CSL for filamentous processes, where it is mainly used as N-source. For this reason, the kinetic release was evaluated especially for the amino acids in CSL.

Bioavailability of amino acids in the solid and liquid phase of CSL

With the established methods, over 80% of the CSL composition could be decrypted, indicated as mg g^{-1} CSL. Although the complete CSL including liquid and solid phase was taken into account, no conclusions about the actual bioavailability can be

Table 5.1: CSL of standard quality was analyzed for its major ingredients in order to get an overview of its composition. The substance classes that were analyzed are amino acids, organic acids, reducing sugars, vitamins, trace elements and minerals.

Ala	Arg	Asp/Asn	Cys-Cys	Glu/Gln	Gly	His	Hyp	Ile	Leu
mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}
35.7	44.3	14.6	11.3	42.0	16.0	6.8	0.8	7.4	21.8
Lys	**Met**	**Phe**	**Pro**	**Ser**	**Thr**	**Trp**	**Tyr**	**Val**	
mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}	
5.4	11.9	6.7	27.6	5.0	7.6	0.4	5.8	15.6	
Ace	**For**	**Lac**	**Succ**	**Oxa**					
mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}					
21.5	26.4	307.9	23.0	4.5					
Arabitol	**Arabinose**	**Galactose**	**Glucosamine**	**Glucose**	**Glycerol**	**Mannitol**	**Sorbitol**		
mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}		
0.9	4.5	2.2	1.4	16.5	14.3	2.7	2.1		
Biotin	**B12**	**Niacinamide**	**Nicotinic acid**	**Pantothenic acid**	**Pyridoxine**	**Thiamine**			
mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}			
16.5	0.2	2.4	0.1	0	0.8	4.2			
Na	**K**	**Mg**	**P**	**S**	**Ca**	**Zn**	**Fe**	**Mn**	**Cu**
mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}	mg g^{-1}	µg g^{-1}	µg g^{-1}	µg g^{-1}	µg g^{-1}	µg g^{-1}
2.9	28.3	6.4	27.0	6.3	361.0	105.3	110.8	35.6	5.6

Ni	Se	Co	Mo
µg g^{-1}	µg g^{-1}	µg g^{-1}	µg g^{-1}
1.05	0	0	0

drawn. Therefore, the initial bioavailability of all amino acids was evaluated, i.e. the soluble amino acids in the liquid as well as in the solid phase. Additionally, the amino acid concentration was determined after chemical and enzymatic extraction in the complete CSL, the solid and the liquid phase. The measured amounts per g CSL for the different phases were in accordance and therefore allowed the detailed separation of each of the phases in bound, i.e. not initially bioavailable, and free, i.e. initially bioavailable, amounts of each amino acid (figure 5.2). Phenylalanine, alanine, leucine and isoleucine occur almost only in their free form with less than 10% bound in proteins. On the contrary, over 90% of hydroxyproline, arginine, histidine, glutamic acid, tyrosine and lysine are not initially bioavailable as they are bound in proteins or peptides. The direct precursor amino acids for penicillin production with *Penicillium chrysogenum* are located in the middle, with about 27% of bioavailable cystine and 55% of bioavailable valine (Jørgensen et al. 1995).

The initial bioavailability of the single amino acids additionally depends on the saturation effects of the substances. Hence, the amount of free amino acids could increase in the fermentation media. In order to estimate this effect, $m_{SR,i}$ was investigated by simple solubilization experiments in batch medium. In summary, 16 samples were taken over 166 h and analyzed via HPLC. For each amino acid a linear regression was fitted through the received concentration analyzed at each sample point. The slope was statistically analyzed via the p-value in order to evaluate a significant slope, i.e. a significant change in concentration that should be considered. Statistically, 12 amino acids, namely glutamic acid, phenylalanine, proline, serine, alanine, valine, leucine, threonine, asparagine, isoleucine, aspartic acid and tyrosine, showed a significant slope which indicates a considerable solubilization (figure 5.3). Nevertheless, it could be observed that the solubilization of amino acids takes place only in the first 72 h. The amino acids showing the highest solubilization effect are mostly amino acids that occur mostly in their bioavailable form in CSL anyway, such as alanine,

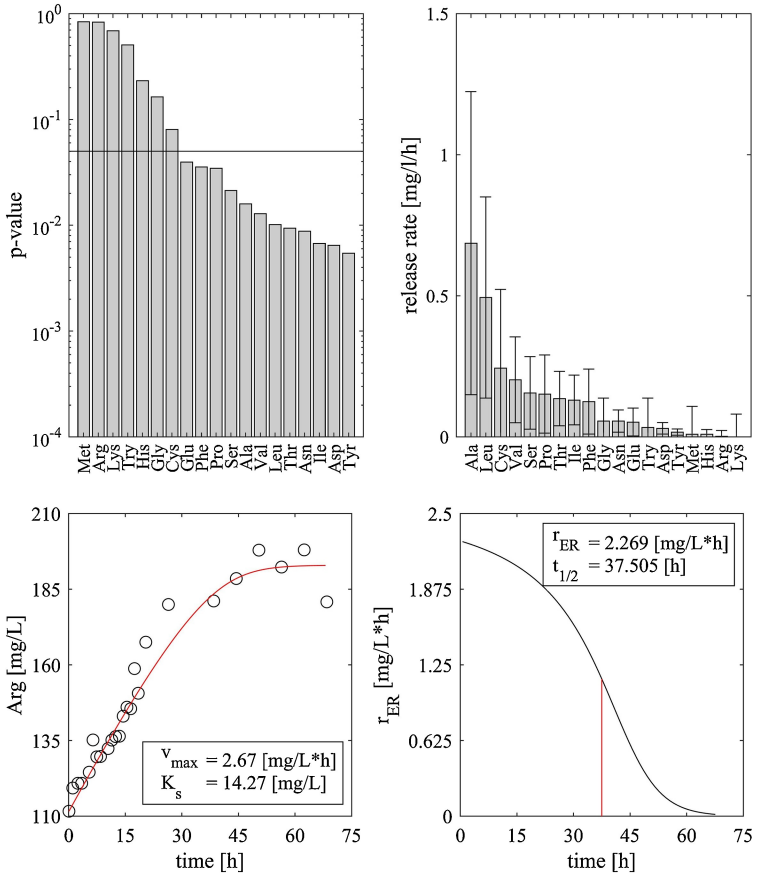

Figure 5.3: Estimation of the solubilization of bound amino acids in media. A regression over the 16 samples was fitted and the p-value of the slope (left plot) was taken as criteria for a significant release. Additionally, the release rate of the single amino acids was calculated (right plot). Twelve out of 19 amino acids showed a significant release.

Figure 5.4: Example of the Monod fit on the example of arginine, including the assumption of bioavailable parts from CSL. The left plot shows the measured concentrations of arginine (white circles), the model fit (red line) and the parameters v_{max} and K_S. The right plot illustrates the release rate r_{ER} over time. The halflife of the r_{ER} is marked by a red line. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

leucine, phenylalanine or isoleucine.

The obtained results raise the question whether the fungi are able to release amino acids from peptides and polypeptides in CSL and whether they are able to metabolize them for growth or production purposes.

Enzymatic release of amino acids from CSL by *P. chrysogenum*

The kinetics of the amino acid release by fungal enzymes was evaluated through enzymes produced in batch and enzymes produced in fed-batch by *P. chrysogenum*. 18 amino acids were investigated for their enzymatic release by peptidases or proteases produced by *P. chrysogenum*. Overall, 10 amino acids showed a marked release, namely arginine, aspartic acid, cystine, glutamic acid, histidine, lysine, serine, threonine, tyrosine and valine. For alanine, glycine, isoleucine, leucine, methionine, phenylalanine, proline and tryptophan no marked release could be noticed. Four of these eight amino acids that show no enzymatic release are available as free amino acids with about 90% of their overall content in CSL. In addition, they also present a solubilization effect as shown above. The bound parts of glycine, methionine and tryptophan on the other hand seem to occur in an unavailable form for *P. chrysogenum* as they show neither solubilization effects nor any enzymatic release.

In order to calculate the release rates $r_{ER,i}$ for the single amino acids, a Monod model was fitted to each run. It is noted here that the assumption of bioavailable parts from CSL had to be included

Table 5.2: The enzymatic release rates of the amino acids available in CSL were evaluated experimentally with *P. chrysogenum*. 10 amino acids showed a marked enzymatic release. The volumetric enzymatic release rates and their half-life of the single amino acids from CSL were evaluated by a Monod model.

	Arg	Asp	Cys-Cys	Glu	His	Lys	Ser	Thr	Tyr	Val
r_{ER} [mg L^{-1} h^{-1}] [a]	2.476	0.699	1.659	4.361	1.080	/	1.890	2.393	0.618	/
r_{ER} [mg L^{-1} h^{-1}] [b]	2.269	0.771	/	3.668	1.299	1.522	1.102	1.495	0.882	1.420
$t_{1/2}$ [h] [a]	30.302	16.486	29.083	18.047	30.132	/	15.993	15.824	23.206	/
$t_{1/2}$ [h] [a]	37.502	13.282	/	22.899	23.154	13.990	27.496	21.363	13.721	18.200

[a] results of experiment 1.
[b] results of experiment 2.

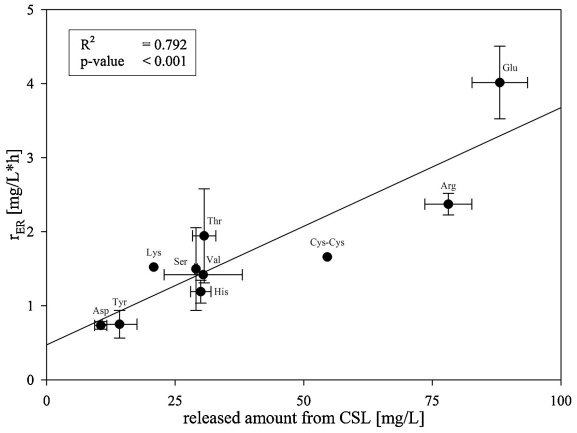

Figure 5.5: The plot shows the released amount of i from CSL against the enzymatic release rate r_{ER} of i for all 10 amino acids that are affected by the fungal enzymes. A correlation with an R^2 of 0.792 can be observed indicating a decrease in the enzymatic activity over time.

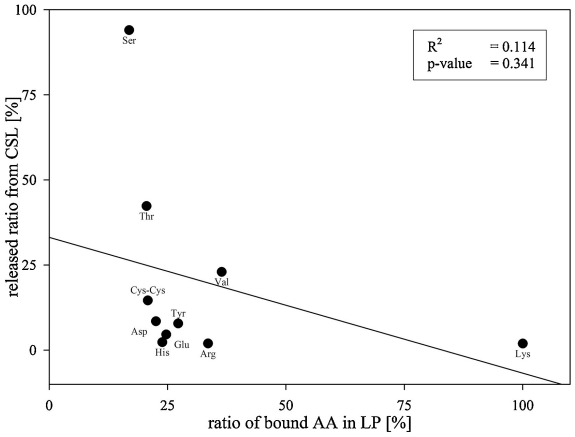

Figure 5.6: The plot shows the ratio of released substance from CSL against the ratio of bound substance in the liquid phase for all 10 amino acids that are enzymatically released by the fungi. No significant correlation can be observed, which indicates that the enzymatic release is independent of the phase.

in the model in order to fit the data. Additionally, no parameter identification was performed for v_{max} and K_s. Hence, these parameters are not reliable. Nevertheless, release rates as well as the half-life of the release rate can be calculated from the data (figure 5.4). A summary of the release rates for both runs and the half-life values for all amino acids that show considerable release are listed in table 5.2.

The enzymatic release rates vary between 0.6 and 4.3 mg L^{-1} h^{-1} for the ten amino acids. Although the amount of bound substance varies for every amino acid, no correlation between the amount of bound amino acid i and r$r_{ER,i}$ could be observed. However, a significant correlation between the amount of released substance i and $r_{ER,i}$ is noticeable (figure 5.5) This correlation as well as the decrease of the release rate over time can arise due to either a decrease in the enzymatic activity over time or due to a limited amount of bioavailable substance in the bound phase. We assume that the first assumption is correct.

As CSL cannot only be divided into free and bound amino acids but also in its liquid and solid phase, the question arises if the enzymatic release kinetics differ between the various phases. In order to investigate this effect, correlations between the ratio of released substance and the ratio between bound

Table 5.3: CSL is added as media supplement in the batch phase of a *P. chrysogenum* process. The amino acids show different behavior in the batch: i) no enzymatic release at all ii) complete release in batch iii) initiation of enzymatic release in batch or iv) no enzymatic release in the batch phase.

AA without $r_{ER,i}$	AA complete release in batch	AA starting $r_{ER,i}$ in batch	AA no release in batch
Ala, Gly, Ile, Leu, Met, Phe, Pro, Trp	Ser, Thr	Arg, Asp, Cys-Cys, Glu, His, Lys, Tyr	Val

substance in the liquid or solid phase were analyzed. No trend was observable that would indicate a preferred phase for the release, hence, we assume that the release kinetics are independent of the phases (figure 5.6).

In summary, the enzymatic release kinetics of 10 amino acids by *P. chrysogenum* could be identified. The question arises how the release of the single amino acids affects the interpretation of a cultivation experiment.

Application of release kinetics in a bioprocess

The results presented so far indicate that if a multiphase system is added as nutrient source it is difficult to estimate the actual bioavailability of the nutrients without proper raw material analysis. Hence, in order to achieve sound process understanding and insight in the metabolic behavior, the actual bioavailability of substances in the raw material has to be considered.

With the example of a *P. chrysogenum* process for the production of penicillin, the impact of the consideration of the amino acid bioavailability of CSL is demonstrated. We hypothesize that CSL is added in this process as N-source and that the amino acids of CSL play an essential role for the formation of penicillin.

The process consists of two stages: the batch for germination and growth and the fed-batch for production of penicillin. CSL is added in the initial media, hence, just in the batch. The transfer from batch to fed-batch takes place when the main C-sources, namely Sucrose, Glucose and Fructose, are depleted, which correlates with the pH course (figure 5.7A). Hence, the pH can be used as transfer criterion. Interestingly, a correlation between pH and the concentration of amino acids in the supernatant could be established by a partial least squares (PLS) regression model (figure 5.7B). According to the pH as transfer criterion and the growth in the batch – shown as the carbon dioxide transfer rate (CTR) –, we assume that the release-time of amino acids in the batch phase is approximately comparable in every batch and is therefore set to 20 h.

In the batch phase, the amino acids can be divided in 4 groups: i) amino acids that are not enzymatically released at all ii) amino acids that are completely released in the batch phase iii) amino acids starting their enzymatically release in the batch and iv) amino acids that are not yet released in the batch (table 5.3). The first two groups are easy to handle as they need no further attention in the fed-batch phase. Most of the amino acids of these two groups are used by the cell for biomass formation. Only methionine and serine are precursor amino acids for cysteine which is important for penicillin formation. Hence, they might play a role for product formation too (Jørgensen et al. 1995; Van Gulik et al. 2000). The latter two groups demand consideration in the fed-batch phase for proper interpretation of uptake and release rates. These two groups include amino acids important for biomass generation, e.g. glutamic acid, a component that can be fed in the TCA cycle, but also cystine and valine, the major precursor amino acids for penicillin production (Jørgensen et al. 1995; Van Gulik et al. 2000; Christensen et al. 1995). In the penicillin cultivations, CSL is added in the media in amounts around 80 g l^{-1}. At the transfer from batch to fed-batch, a dilution takes place decreasing the amount of transferred CSL. Hence, for some amino acids the enzymatic release in the fed-batch just lasts for the first 10–50 h. In table 5.4 the concentrations of the different amino acids in the different phases is illustrated.

On the example of two amino acids, one important for biomass formation the other for product formation, the impact of the release from CSL is illuminated. When we compare the cumulative specific

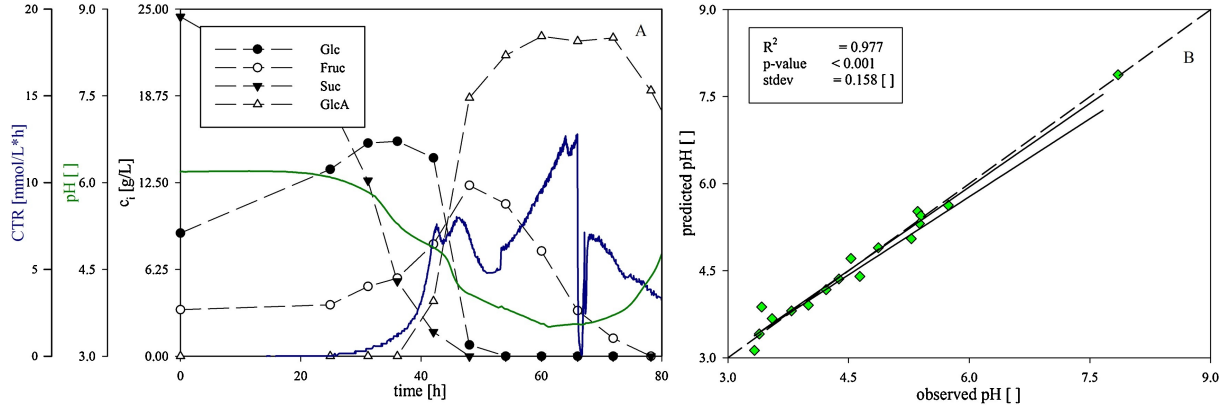

Figure 5.7: Plot A shows a typical time course of sugars, pH and carbon dioxide transfer rate (CTR) – representing growth – in a *P. chrysogenum batch cultivation*. Sucrose is cleaved into glucose and fructose, which are taken up consecutively. With the uptake of glucose, the production of gluconic acid (GlcA) starts, which is jointly responsible for the change in culture pH. The transfer to the fed-batch culture takes places after 75 h when glucose and fructose are completely consumed. This correlates to the culture pH, which can therefore be used as transfer criterion. Plot B shows a PLS model with 5 latent variables of pH and the amino acid concentrations in the supernatant of the batch culture. The good correlation allows an assumption of a constant amino acid release time in the batch if the pH is used as transfer criterion.

Table 5.4: Overview of the concentrations of all amino acids that show an enzymatic release at the beginning of the batch in free form, bound, released during batch and transferred to the fed-batch.

	Arg	Asp	Cys-Cys	Glu	His	Lys	Ser	Thr	Tyr	Val
	mg L^{-1}	mg L^{-1}	mg L^{-1}	mg L^{-1}	mg L^{-1}	mg L^{-1}	mg L^{-1}	mg L^{-1}	mg L^{-1}	mg L^{-1}
Begin batch (incl $m_{ER,i}$)	34.9	41.2	98.0	80.9	9.8	26.1	236.2	142.8	17.2	274.0
Bound in CSL	1551.7	121.7	258.4	1290.4	295.7	224.5	53.3	87.2	212.7	225.1
Released in batch	47.4	14.7	33.2	80.3	23.8	30.4	29.9	38.9	15.0	0
Transferred to fed-batch (bound)	224.5	16.0	33.6	180.6	40.6	29.0	3.5	7.2	29.5	33.6

uptake rate of i with either the cumulative growth rate or production rate we can observe some correlation. This correlation changes, if the release from CSL is taken into account, hence, a different yield value is observed. Arginine is part of the urea cycle and is mainly used as N-source for biomass formation (Morris Jr. 2002; Wu and Morris Jr. 1998). Additionally, it is present in CSL in high quantity and is released from CSL in the fed-batch for about 130 h (Cardinal and Hedrick 1948). Without taking any release into account, we get a positive yield value, assuming a production of arginine by the fungi. Contrariwise, an uptake of arginine and a negative yield can be observed, if the release from CSL is taken into account (figure 5.8A). Although no reference yield for biomass/arginine for fungi in complex medium could be found in literature, the observed correlation is consistent when looking at the metabolic pathway of arginine and its role as nitrogen source for biomass formation (Van Gulik et al. 2000; Morris Jr. 2002; Wu and Morris Jr. 1998). The same approach was performed for valine, a precursor amino acid for penicillin production, and penicillin. Valine is part of the α-aminoadipyl-cysteinyl-valine (ACV) peptide that is converted to isopenicillin N. The sidechain of isopenicillin N is finally exchanged with phenoxyacetate and forms penicillin V (Jørgensen et al. 1995). In the process valine is

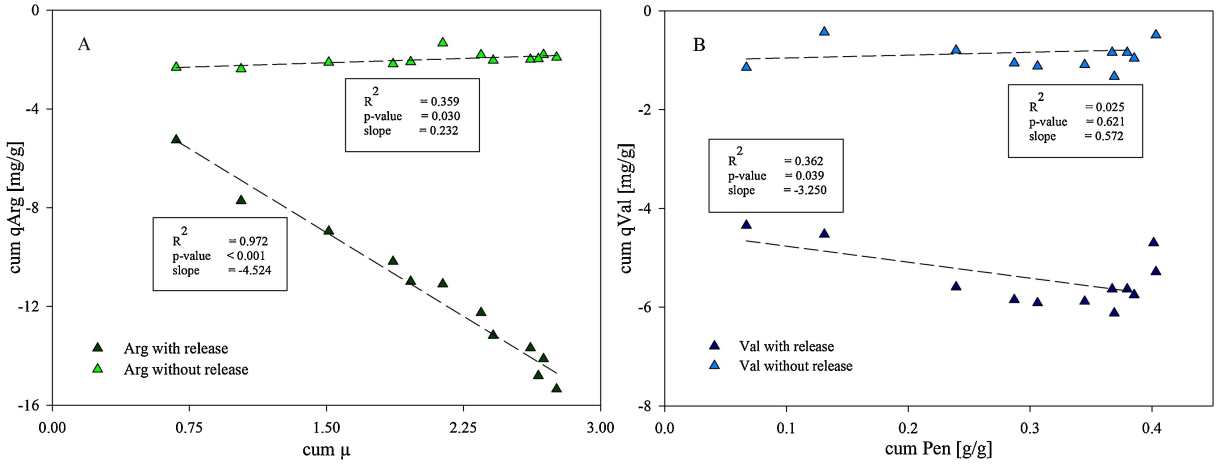

Figure 5.8: Correlation between arginine and biomass (A) as well as valine and penicillin (B). Both plots show the cumulated specific rates of biomass and penicillin versus the cumulated specific rates of arginine and valine, respectively. In both cases a marked difference between these correlations can be observed depending on a consideration of amino acid release from CSL or not, which can also be observed in the slope change which corresponds to the yield coefficient.

just released for the first 32 h in a standard fed-batch. Nevertheless, the observed yield and interpretation of the specific rate of valine change markedly. Without consideration of the release kinetics, the yield is positive and a production of valine by the fungi is indicated. The interpretation changed into the opposite again, when taking the release kinetics of CSL into account (figure 5.8B). Valine is taken up by the fungi and we get a negative yield. As valine is a precursor amino acid, this correlation seems plausible, too.

As mentioned above, the proposed kinetics can have an impact on the interpretation of physiological activity of the fungi. Taking a look at the concentration of arginine in the supernatant, an unsteady behavior can be observed with a peak at approximately 70 h. Additionally, the release from CSL takes about 130 h for arginine and is assumed to be constant (figure 5.9A). The release of bound arginine from CSL already indicates an uptake of this amino acid by the fungi, as the concentration in the supernatant changes just marginally. The specific rates calculated with or without release kinetics substantiated this hypothesis (figure 5.9B). In the first 12 h of the fed-batch an uptake of arginine takes place. Without consideration of any release, the specific arginine rate just varies around zero after this initial uptake. The arginine peak that can be observed in the supernatant is only represented by a sudden release of arginine by the fungi. If the release by CSL is taken into account, a strong initial uptake can be observed, followed by a constant uptake of arginine by the fungi. This difference is furthermore illustrated by the cumulated specific arginine rate over time. The release from CSL results in a steady process constantly taking up arginine (figure 5.9D), whereas no release shows an insignificant trend that indicates no metabolic action by the fungi concerning arginine (figure 5.9C).

Additionally to the change in physiological interpretation, the release kinetics might also help to identify phases in the process. This phenomenon is shown again in a penicillin fed-batch process with less initial CSL concentration. Hence, the release of arginine in the fed-batch lasts for around 60 h (figure 5.10A). Without any consideration of a release a constant specific arginine rate can be observed around zero (figure 5.10B). If the release is taken into account, the process suddenly shows two phases. The first phase till 80 h shows an uptake of arginine

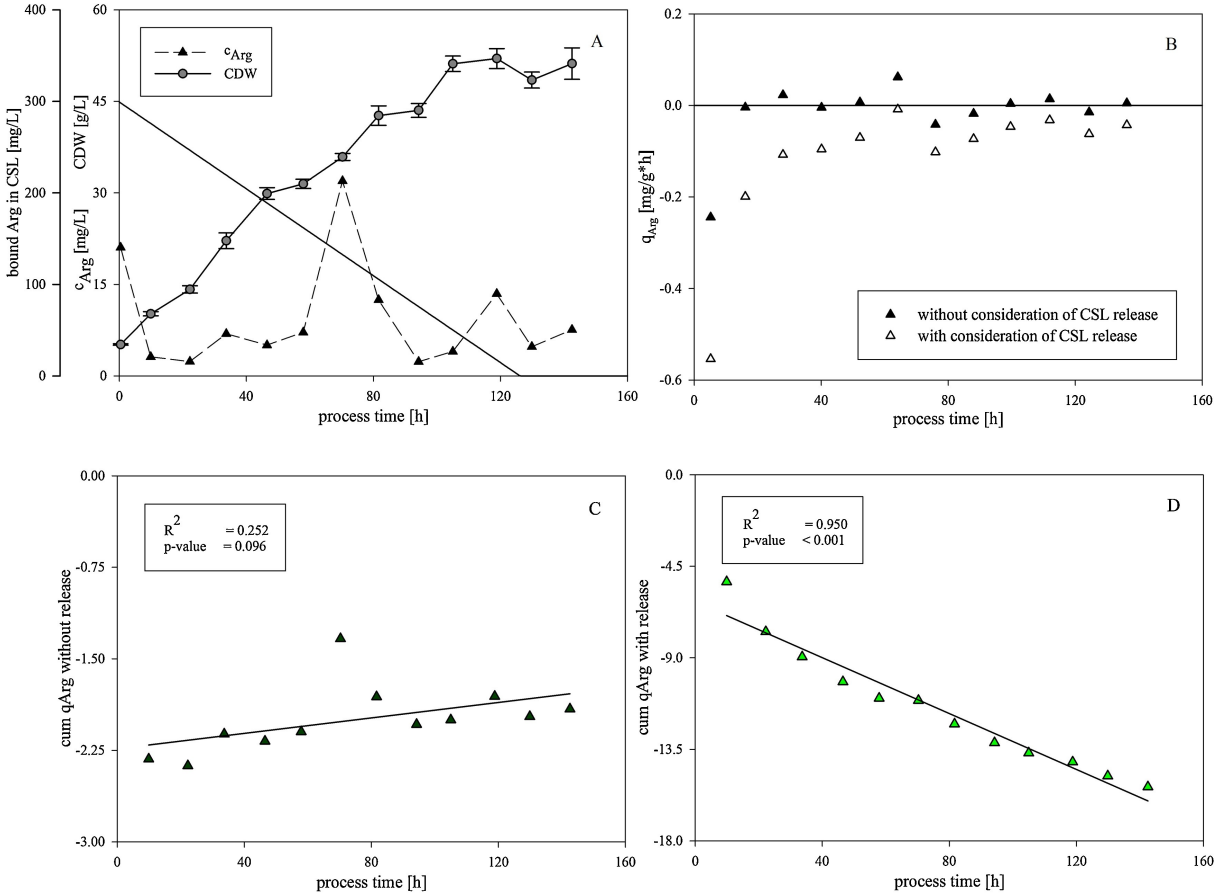

Figure 5.9: Impact on the interpretation of physiological behavior of the fungi with consideration of CSL release in the fed-batch. Plot A shows the time course of biomass and arginine in the supernatant over time as well as the decrease in bound arginine due to an enzymatic release (black line). Plot B illustrates the specific arginine rates calculated with or without release from CSL. The difference in the cumulated specific rates over time can be seen in plot C and D.

that decreases over time. The second phase illustrates a constant but very low release rate of arginine that equals the one without consideration of any enzymatic release from CSL. Again, the interpretation of the physiology of the fungi in the processes changes due to the consideration of enzymatic release of amino acids from CSL. In the first 24 h of this process a C-limitation occurred, which might be the reason for an intensified uptake of amino acids. After removal of the limitation, the arginine uptake rates decreased again because another C-source was again available. The hypothesis about the occurrence of these two phases could now be experimentally investigated and substantiated. This is likely to lead to an increased understanding of the organism as well as of the process itself.

Conclusion

CSL is a widely used raw material in bioprocesses, which still holds secrets as to why it has such a positive impact on process performance. In the process of *Penicillium chrysogenum* it is mainly added as N-source due to its high con-

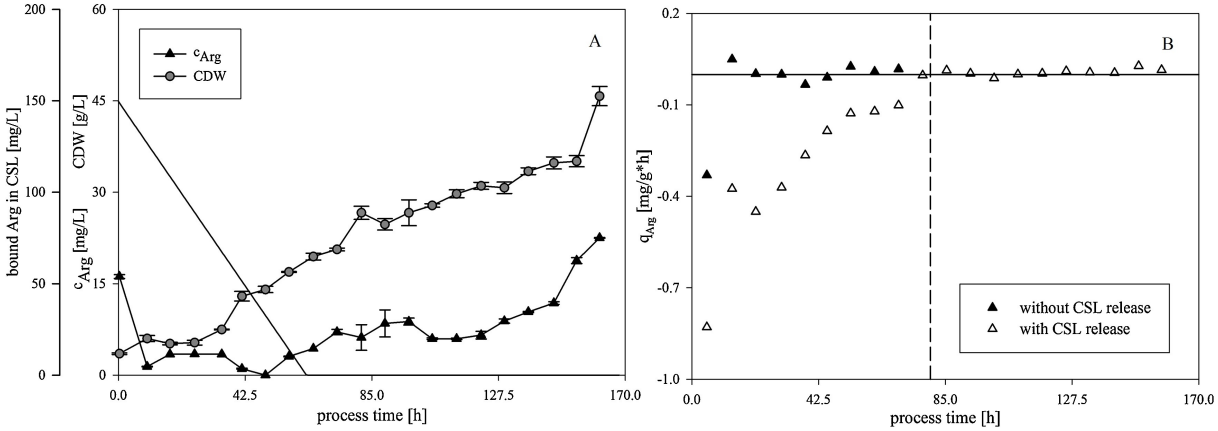

Figure 5.10: The consideration of enzymatic release kinetics from CSL can help to detect phases in a process that can otherwise not be observed that easily. Plot A shows the concentrations of biomass and arginine over time as well as the decreasing concentration of bound arginine in CSL (black line). Plot B illustrates the specific arginine rates that show one phase without release consideration and two phases with release consideration.

tent of amino acids. CSL is a multicomponent and multiphase system, consisting of both bound and free ingredients in its liquid and solid phase. Hence, bioavailability and the actual nutritious effect of the amino acids are not known during bioprocesses.

In this contribution, we give a holistic overview of CSL by characterizing CSL for 55 specific components and specifying both phases, solid and liquid. Initial bioavailability of amino acids in both phases of CSL was determined as well as the amount of bound amino acids. The release rates from the peptides and polypeptides in CSL were determined with respect to i) solubilization effects in media and ii) enzymatic decomposition by enzymes produced from *P. chrysogenum*. It could be shown that these release kinetics play a role in the understanding of a process that uses CSL as complex media supplement. On the example of valine and arginine in the penicillin process, the change in physiological interpretation of release and uptake kinetics of the fungi were illustrated just by consideration of CSL release kinetics. For both amino acids an uptake can be observed if release kinetics are considered, whereas no consideration implies a production of these amino acids by the fungi. Additionally, it could be shown that phase recognition can be facilitated by inclusion of amino acid release from CSL.

We believe that the identification and application of enzymatic release kinetics from CSL are beneficial to all processes that rely on this complex raw material. They will increase the understanding of the cell physiology and the process itself and might therefore be beneficial to process optimization.

Conflict of interest

The authors declare no conflict of interest.

Funding

Financial support was provided by the Austrian research funding association (FFG) under the scope of the COMET program within the research project "Industrial Methods for Process Analytical Chemistry – From Measurement Technologies to Information Systems (imPACts)" (contract # 843546).

Acknowledgements

The authors would like to thank Sandoz GmbH for providing the strain. Additionally, we want to thank Winfried Nischkauer for performing the ICP-OES measurements. We want to thank Katharina Oberhuber for English proof reading.

References

Bowden, J. P. and W. H. Peterson (1946). "The role of corn steep liquor in the production of penicillin." In: *Archives of biochemistry* 9, p. 387.

Cardinal, E. V. and L. R. Hedrick (1948). "Microbiological assay of corn steep liquor for amino acid content". In: *Journal of Biological Chemistry* 172.2, pp. 609–612.

Christensen, L. H. et al. (1995). "Continuous cultivation of *Penicillium chrysogenum*. Growth on glucose and penicillin production". In: *Journal of biotechnology* 42.2, pp. 95–107. DOI: 10.1016/0168-1656(95)00056-V.

FDA (2004). "Guidance for industry, PAT-A Framework for Innovative Pharmaceutical Development, Manufacturing and Quality Assurance". In: *www.fda.gov*.

Gao, Y. and Y. Yuan (2011). "Comprehensive quality evaluation of corn steep liquor in 2-keto-l-gulonic acid fermentation". In: *Journal of agricultural and food chemistry* 59.18, pp. 9845–9853. DOI: 10.1021/jf201792u.

Hofer, A. and C. Herwig (2017). "Quantitative determination of nine water-soluble vitamins in the complex matrix of corn steep liquor for raw material quality assessment". In: *Journal of Chemical Technology & Biotechnology* 92.8, pp. 2106–2113. DOI: 10.1002/jctb.5211.

Hofer, A., D. Kamravamanesh, et al. (2018). "Prediction of filamentous process performance attributes by CSL quality assessment using mid-infrared spectroscopy and chemometrics". In: *Journal of biotechnology* 265, pp. 93–100. DOI: 10.1016/j.jbiotec.2017.11.010.

Hull, S. R. et al. (1996). "Composition of corn steep water during steeping". In: *Journal of agricultural and food chemistry* 44.7, pp. 1857–1863. DOI: 10.1021/jf950353v.

Jørgensen, H. et al. (1995). "Metabolic flux distributions in *Penicillium chrysogenum* during fed-batch cultivations". In: *Biotechnology and Bioengineering* 46.2, pp. 117–131. DOI: 10.1002/bit.260460205.

Kanzaki, T. et al. (1967). "l-Glutamic Acid Fermentation: Part I. Selection of an Oleic Acid-Requiring Mutant and its Properties Part II. The Production of l-Glutamic Acid by an Oleic Acid-Requiring Mutant". In: *Agricultural and Biological Chemistry* 31.11, pp. 1307–1317. DOI: 10.1080/00021369.1967.10858972.

Kona, R. P., N. Qureshi, and J. S. Pai (2001). "Production of glucose oxidase using *Aspergillus niger* and corn steep liquor". In: *Bioresource technology* 78.2, pp. 123–126. DOI: 10.1016/S0960-8524(01)00014-1.

Liggett, R. W. and H. Koffler (1948). "Corn steep liquor in microbiology". In: *Bacteriological reviews* 12.4, p. 297.

Morris Jr., S. M. (2002). "Regulation of enzymes of the urea cycle and arginine metabolism". In: *Annual review of nutrition* 22.1, pp. 87–105. DOI: 10.1146/annurev.nutr.22.110801.140547.

Posch, A. E. et al. (2013). "Combining light microscopy, dielectric spectroscopy, MALDI intact cell mass spectrometry, FTIR spectromicroscopy and multivariate data mining for morphological and physiological bioprocess characterization of filamentous organisms". In: *Fungal genetics and biology* 51, pp. 1–11. DOI: 10.1016/j.fgb.2012.11.008.

Rivas, B. et al. (2004). "Development of culture media containing spent yeast cells of *Debaryomyces hansenii* and corn steep liquor for lactic acid production with *Lactobacillus rhamnosus*". In: *International journal of food microbiology* 97.1, pp. 93–98. DOI: 10.1016/j.ijfoodmicro.2004.05.006.

Silveira, M. et al. (2001). "Production of glucose-fructose oxidoreductase and ethanol by *Zymomonas mobilis* ATCC 29191 in medium containing corn steep liquor as a source of vitamins". In: *Applied microbiology and biotechnology* 55.4, pp. 442–445. DOI: 10.1007/s002530000569.

Tornqvist, E. G. M. and W. H. Peterson (1956). "Penicillin production by high-yielding strains of *Penicillium chrysogenum*". In: *Applied microbiology* 4.5, p. 277.

Van Gulik, W. M. et al. (2000). "Application of metabolic flux analysis for the identification of metabolic bottlenecks in the biosynthesis of penicillin-G". In: *Biotechnology and Bioengineering* 68.6, pp. 602–618. DOI: 10.1002/(SICI)1097-0290(20000620)68:6%3C602::AID-BIT3%3E3.0.CO;2-2.

Wu, G. and S. M. Morris Jr. (1998). "Arginine metabolism: nitric oxide and beyond". In: *Biochemical Journal* 336.1, pp. 1–17. DOI: 10 . 1042 / bj3360001.

Xiao, X. et al. (2013). "Classification and analysis of corn steep liquor by UPLC/Q-TOF MS and HPLC". In: *Talanta* 107, pp. 344–348. DOI: 10.1016/ j.talanta.2013.01.044.

Metabolic flux analysis linked to complex raw materials as tool for bioprocess improvement

This manuscript was published in *Chemical Engineering Science*.

Citation: Hofer, A., **Hauer, S.**, Kroll, P., & Herwig, C. (2018). Metabolic flux analysis linked to complex raw materials as tool for bioprocess improvement. Chemical Engineering Science, 191, 245-252.

Submitted: 3rd April 2018

Available online: 28th June 2018

DOI: 10.1016/j.ces.2018.06.075

My contribution: Set up the stoichiometric model from literature models and KEGG data. Performed the experiments together with Alexandra Hofer and Julian Kager. Calculated rates from data and performed metabolic flux analysis with sensitivity analysis.

Metabolic flux analysis linked to complex raw materials as tool for bioprocess improvement

Alexandra Hofer[a], Stefan Hauer[a], Paul Kroll[a,b], Christoph Herwig[a,b]

[a] Research Area Biochemical Engineering, Institute of Chemical, Environmental and Bioscience Engineering, Vienna University of Technology, Gumpendorferstrasse 1a – 166/4, A-1060 Vienna, Austria
[b] CD Laboratory on Mechanistic and Physiological Methods for Improved Bioprocesses, Vienna University of Technology, Gumpendorferstrasse 1a – 166/4, A-1060 Vienna, Austria

Abstract

BIOPHARMACEUTICAL processes should be designed to maximize productivity and ensure product quality. What is underestimated in this context is the variability of the raw material. Especially complex raw materials are challenging to characterize, hence, the identification of crucial raw material attributes that influence productivity or product quality is troublesome. In this study metabolic flux analysis (MFA) is applied to fill the gap between raw material characterization and process understanding. The approach is demonstrated in corn steep liquor (CSL) and a *Penicillium chrysogenum* process. This bioprocess is already well understood and there exist various MFA models aiming at understanding the fungal metabolism and production pathways. However, MFA was as per our knowledge not used for the identification of critical raw material attributes. Hence, in this study an MFA model was adapted from literature including CSL related fluxes and CSL release kinetics. The application of a sensitivity analysis with respect to q_{Pen} and μ, revealed the potential of the model based approach: we identified methionine as a key attribute in CSL for penicillin production. As a consequence, an optimized process could be presented by reducing CSL in the media and pulsing methionine, which resulted in a duplication of product titer.

In summary, the expansion of an MFA model with raw material characteristics featured by the application of sensitivity analysis is a promising approach for science-based decisions on crucial raw material attributes. It could facilitate the predictive design of complex raw materials along Quality by Design rationales as well as model-based process improvement with respect to raw material attributes. Additionally, the method allows the identification of raw material variability and the impact of these variances on the process.

Introduction

EVERY biopharmaceutical production process aims for high productivity by maintaining the predetermined product quality. In order to guarantee product quality, regulatory guidelines demand sound science-based process understanding (Guideline 2009). This includes the knowledge about functional relationships between critical material attributes (CMAs) or critical process parameters (CPPs) and critical quality attributes (CQAs). The mechanism between process parameters and productivity is defined by key process parameters (kPPs). In terms of media development and complex raw material, also material attributes may influence productivity, hereby called key material attributes (kMAs). The approach devised to achieve a process design by science-based decisions based on the knowledge of functional relationships between inputs and output is called Quality by Design (QbD).

Especially for complex raw material, the QbD approach is challenging to follow (Lanan 2009). E.g. for corn steep liquor (CSL), the identification of CMAs or kMAs is difficult as it is a multicomponent and multi-phase mixture delivering over 55 specific compounds (Hofer et al. 2018). Additionally, it underlies a high lot-to-lot variability due to its biological nature and shows specific characteristics concerning bioavailability during a process (Liggett and Koffler 1948). Recently, CSL has been characterized in detail including an investigation of the bioavail-

ability of amino acids during processes with *Penicillium chrysogenum* (unpublished results). In this process, CSL is mainly added as N-source making use of the 21 amino acids that occur in bound and free form (Pyle 1954).

The process for the production of penicillin is very well studied. Various contributions have been published that used metabolic flux analysis (MFA) for understanding the fungal metabolism (Agren et al. 2013; Christensen and Nielsen 2000) and the penicillin production pathway (Jørgensen et al. 1995; Kleijn et al. 2007; Van Gulik et al. 2000). Penicillin production requires an ACV peptide formed of α-amino adipic acid, cysteine and valine, and a precursor for the side chain addition, namely phenoxyacetate for the production of penicillin V. As CSL serves as amino acid source, including all three amino acids necessary for the ACV peptide, the usage of a MFA model for linking raw material quality with process performance is obvious. Model-based methods are commonly applied in biopharmaceutical industry for the generation and storage of knowledge (Kroll et al. 2017). However, to the author's knowledge no approach has been presented so far combining complex raw material analytics with MFA modeling.

In this study, an MFA model is adapted from literature and expanded by CSL specific characteristics. This means an inclusion of all amino acid fluxes as well as enzymatic release kinetics from CSL for amino acid rate calculation. We show a novel approach that an MFA model can be combined with sensitivity analysis in order to generate hypotheses about potential kMAs. A limiting amino acid was detected and by changing the media and process control/conduction the product titer could be doubled. Hence, by introducing a sensitivity analysis, limitations from CSL could be identified and MFA could be used for process improvement.

The presented approach of combining MFA and sensitivity analysis can be the basis for a science-based approach of raw material quality assessment and the identification of CMAs and kMAs as demanded by regulatory guidelines. It is a step towards QbD for complex raw materials in biopharmaceutical production.

Materials and methods

Chemicals and reagents

All chemicals and substances were of analytical grade and purchased from either Carl Roth (Karlsruhe, Germany) or Sigma Aldrich (St. Louis, MO, USA). For analytical methods ultra-pure water was used, derived from a Milli-Q system from Merck Millipore (Billerica, MA, USA).

Analytical methods

Amino acids, sugars and sugar alcohols were analyzed according to Hofer et al. (2018).

Penicillin V and phenoxyacetate (Pox) were quantified by HPLC. Chromatographic separation was achieved by a Zorbax Eclipse AAA column, 4.6 $\times$ 12.5 mm, 5 µm (Agilent Technologies, California, USA) with an eluent composed of 28% acetonitrile, 0.005 M potassium dihydrogen phosphate and 0.07% phosphoric acid and a flow rate of 1.0 mL min^{-1}. The column oven temperature was set to 30 °C and the analytes were detected at 192 nm.

Dry cell weight (CDW) was determined by centrifugation of 5 mL cultivation broth at 4800 rpm, 4 °C for 10 min. The pellet was resuspended once in deionized water and subsequently centrifuged and the cell pellet dried at 105 °C for 72 h.

Bioreactor cultivations

All cultivations were performed with a strain of *Penicillium chrysogenum* as used by Posch, Koch, et al. (2013). Cultivations were performed in a DASGIP multi-bioreactor system (Eppendorf AG, Germany) with a working volume of 2 L each. The initial batch and fed-batch volumes were 2 L and 1.5 L, respectively. Each reactor was equipped with optical and potentiometric probes (Hamilton, Switzerland) for measuring dissolved oxygen (dO_2) and pH, respectively. dO_2 was controlled via adjustment of stirrer speed and increase of the O_2 content in the inflow air. Aeration was kept at 1 vvm. pH was controlled via addition of 20% (w/v) KOH or 15% (v/v) H_2SO_4. CO_2 and O_2 concentrations were analyzed using an off-gas analyzer (module DASGIP GA4) by a ZrO_2

sensor for O_2 and an infrared sensor for CO_2 analysis.

The batch was carried out on a complex medium similar to the one described by Posch, Koch, et al. (2013). Standard batch medium included 80 g L^{-1} CSL as single complex ingredient. During batch cultivation dO_2 and temperature were controlled over 40% and at 25 °C, respectively. pH was not controlled as the end of the batch was indicated by an increase in pH by 0.5. Fed-batch was performed on a defined medium (similar to Posch, Koch, et al. (2013)), which was inoculated by 10% of the batch broth. During fedbatch cultivation pH (6.5), dO_2 (¿40%) and temperature (25 °C) were controlled. Three feeds were added to the fed-batch, namely glucose, ammonia and a pox feed, via peristaltic pumps (module DASGIP M8, Eppendorf AG, Germany).

Metabolic flux model

For metabolic flux balancing, the well established approach described by Goudar et al. (2014) was used as depicted in Eq. (1):

$$A \times \vec{v} = 0 + \vec{\epsilon} \tag{5.1}$$

With A being the bioreaction network containing all fluxes including the exchange reactions between media and cytosol, i.e., uptake and production fluxes, $\vec{v}$ being the flux vector of all intracellular and exchange fluxes and $\vec{\epsilon}$ being the error. Within this approach a data reconciliation is performed, as described by equation 16 in the same publication. The reconciliation leads to the most probable specific rates, taking into account redundancies and pre-examined measurement errors. The challenge here is the estimation of the errors on the specific rates. For this a Monte- Carlo approach was chosen: Errors on variables, i.e., concentrations and specific rates, where considered to be normally distributed. Using this assumption, a set of n = 200 states were used to perform calculations in order to estimate the specific rates and there standard deviation. The mean and standard deviation of this simulated set of rates were used as inputs for metabolic flux balancing. All calculations were done using Matlab version R2016b (Mathworks Inc., USA).

Rate calculation

Rate calculation was performed as described in Kroll et al. (2017).

Sensitivity analysis

For sensitivity analysis the response flux of interest was treated as an unknown flux and all fluxes were then iteratively deflected by a set increment. The unknown fluxes were then recalculated with the changed inputs and the sensitivity was defined as the first order derivative, i.e., the change in output flux divided by the deflection of the input flux. This is equivalent to the differential analysis approach reviewed by Sin, Gernaey, and Lantz (2009).

Results

Adapted MFA model for the *P. Chrysogenum* process for penicillin V production

There are multiple metabolic flux models available in literature for *Penicillium chrysogenum* (Agren et al. 2013; Christensen and Nielsen 2000; Jørgensen et al. 1995; Kleijn et al. 2007; Van Gulik et al. 2000). This study focuses on the functional relationships between complex raw material and process performance. Hence, a model was adapted from several literature models (Henriksen et al. 1996; Jørgensen et al. 1995; Van Gulik et al. 2000), and further adjusted and simplified using the Kyoto encyclopedia of genes and genomes. The model is unsegregated and no compartments inside the cell are taken into account. It includes glycolysis, the pentose phosphate pathway, the TCA cycle, the mannitol cycle, oxidative phosphorylation, the amino acid metabolism, biomass formation and penicillin biosynthesis. As CSL functions as N-source in this process, all 18 amino acids are included in the model. In total 76 fluxes are included, which are all listed in the Appendix Table A1. Biomass was set according to Van Gulik et al. (2000). The specific rate calculation for

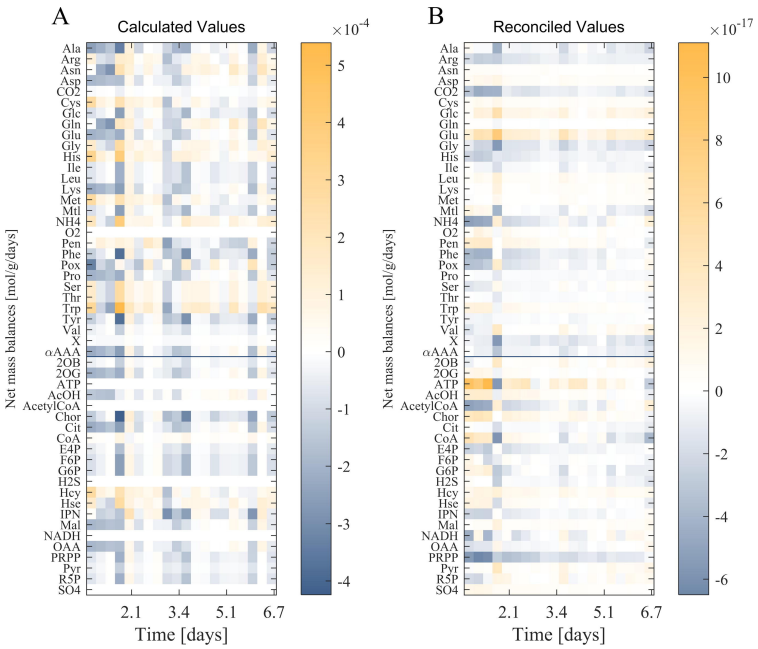

Figure 5.11: Net mass balances of the MFA model of the measured components (above the horizontal black line) and the unmeasured components (under the horizontal black line) over the whole process. Assuming a steady state at every point of time in the process the net mass balances should theoretically be zero, but actually resemble $\vec{\epsilon}$. Plot A shows the actual net mass balances for the MFA model using all measured states as input for estimating the unknown fluxes, while plot B uses the reconciled rates as input. The net mass balances could be improved by a factor of 10^{14} with the reconciled values (plot B).

amino acids was adapted to the previously evaluated enzymatic release kinetics from CSL. The final model includes 51 metabolites, 76 fluxes, 21 pathways and 29 measured fluxes. It shows a full matrix rank, a condition number of 160 and a degree of freedom of -4.

For quality assessment, the model was applied on a standard *P. chrysogenum* cultivation run and balances were evaluated. The applied MFA model is a static model that assumes a steady state at every time point in the process. Hence, the net mass balances for every component should be zero. The actual values for all measured and estimated components were evaluated and resulted in values between -4 $\times$ 10^{-4} and 5 $\times$ 10^{-4} (figure 5.11). Due to redundancy and evaluation of measurement errors by Monte Carlo simulations, data reconciliation could be performed. Reconciliation could significantly improve the net mass balances by the factor 1014 (note the scaling in figure 5.11). It should be noted that the balances close for sulfur related components, whereas they show most deviations from zero for ATP, NH_4, CO_2, Acetyl-CoA and PRPP, which decrease over process time. Additionally, an estimation of reliable measurements can be made. The data seem to show an outlier measurement after around 48 h. The reconciled data were used for further application of the MFA model.

Sensitivity analysis for detection of limitations

MFA models are commonly applied to get an insight into the metabolism of the organism, e.g. to detect energetic bottlenecks or to visualize the precursor substances for a product. In this case, we want to generate knowledge of the correlations between CSL composition and fungal metabolism. Hence, we hypothesize that the application of a sensitivity analysis on the MFA model, i.e. a random deflection of the fluxes, can visualize important raw material ingredients on process performance. Targets of the sensitivity analysis are the specific production rate q_{Pen} and the growth rate μ. Differential analysis was chosen as sensitivity method. The fluxes for q_{Pen} and μ were set to unmeasured simultaneously in order to prevent cross correlated effects of both fluxes. All remaining measured fluxes are randomly deflected 200 times and the fluxes for q_{Pen} and μ are estimated

by the model. Linear regression is subsequently applied to determine the sensitivity of the single fluxes on the targets.

For the sensitivity analysis, a local approach assuming a linear effect was applied. Therefore the deflection of the specific rates was set constant on the value of 10^{-16}. Afterwards the sensitivity was calculated according to equation 5.2, where δy resembles the change in the specific rates after deflection of the input rates (δx).

$$s_{x,y} = \frac{\delta y}{\delta x} \tag{5.2}$$

It should be noted that the size of deflection should be carefully chosen according to the rates used and whether measured or reconciled flux values were used. If deflection is too high, the response will only be dependent on each individual flux because of nonlinear effects, if the deflection is too low, the sensitivity shows only dependency on time because of calculation limits of the used software.

In order to get an overview of the important fluxes, the combination of MFA and sensitivity analysis was applied on a standard run of *P. chrysogenum* (figure 5.12). Please note that a negative effect in this case means that an increase in this flux increases penicillin production, while a positive effect means that an increase in this flux yields a lower penicillin flux. This is due to the different signs of the response and independent fluxes – production fluxes are positive, as uptake fluxes are by definition negative.

Existing knowledge of the process can be consulted in order to evaluate the meaning of the acquired results. Douma et al. (2010) established the correlation between μ and q_{Pen} in fungal processes, implying that high growth rates result in low specific production rates. This correlation can also be seen in the results of the sensitivity analysis and confirms our modeling results. Fluxes of substances that are growth correlated such as q_{Glu} or q_{Gln} show opposite sensitivity on μ and q_{Pen}.

The highest impact on specific productivity can be observed for cystine and methionine, both fluxes indicating a positive correlation with q_{Pen}. The correlation with cystine is not surprising as cysteine is one of the three precursor amino acids of the ACV peptide (Jørgensen et al. 1995). Interestingly, the effect of the other two precursor amino acids, namely valine and a-amino adipic acid, seems negligible. An explanation for this can be seen in the metabolic fluxes. a-amino adipic acid is not essential as it is an intermediate of the lysine pathway (Hönlinger et al. 1988); valine can be generated from pyruvate.

The critical path in penicillin production seems to be the availability of sulfur for the β-lactam ring. Sulfur can be provided in three ways: by the uptake of (i) inorganic sulfate (ii) cysteine or (iii) methionine (Nielsen 1997). Inorganic sulfate can be taken up by an active transport system that is feed-back inhibited by intracellular methionine or cysteine. *P. chrysogenum* also shows specific amino acid transport systems for methionine and cysteine. Upon sulfur-starvation, the specific methionine transporter is developed, which is also feed-back inhibited by intracellular sulfur or sulfur containing compounds. In the exemplary process of this study, inorganic sulfate is not taken up as this approach is energetically expensive for the cell. In order to take up sulfate and convert it to sulfide, which is necessary for the biosynthesis of cysteine, 4 mol NADPH and 2 mol ATP are required (see Appendix Table A1, V39). Hence, two sulfur sources for penicillin production remain, namely cysteine and methionine, which are both provided by CSL in the batch phase.

In order to evaluate the impact of both critical substances two standard Monod models were set up and evaluated with historical datasets. The two models hypothesize that q_{Pen} is directly correlated with c_{Met} or c_{Cys}, respectively. The model stating methionine as the limiting substrate for penicillin production resulted in reasonable model parameters that could both be identified. It can be clearly seen that methionine is completely consumed and that the penicillin production decreases, if no methionine is available any more. In the other case, no reasonable model could be generated for cysteine as the limiting substrate, probably connected to the high measurement error of this substance and the fact that cysteine is not completely depleted during the process (figure 5.13).

In summary, the results of the sensitivity analysis seem reasonable and indicate sulfur as the bottle-

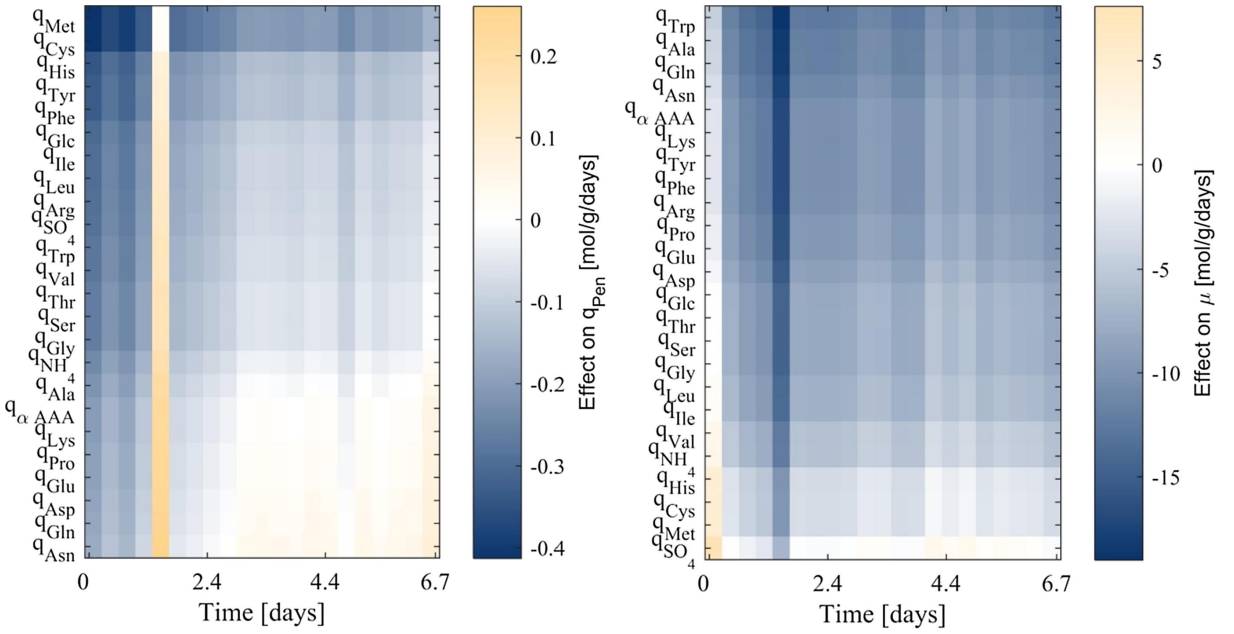

Figure 5.12: Results of the combination of the MFA model and sensitivity analysis. The target responses were q_{Pen} and m, which were both set to unmeasured. Deflection of the other fluxes showed the effect of single rates on qpen (plot A) or m (plot B). A negative effect (blue) in this case means that an increase in this flux will also increase penicillin production or biomass formation, respectively. In plot A q_{Met} and q_{Cys} can be identified as the substances showing the highest potential for an increase in specific productivity. Plot B depicts the impact of the fluxes on m. It can be seen that the sulfur fluxes that impact productivity positively, show an opposite behavior on growth. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

neck for penicillin production. Interestingly, methionine seems to be the limiting substrate according to the MFA results combined with historical data.

Process improvement considering MFA model

The combination of the MFA model and the sensitivity analysis resulted in a hypothesis of methionine being the limiting substrate from CSL media. For experimental substantiation, an experiment was performed with the aim of reaching this limitation at an early stage of the cultivation. Hence, the initial CSL concentration in the batch media was set to 40 g L^{-1}, which is half of the standard concentration. As soon as the limitation was reached, detected by at-line measurements of penicillin V and amino acids, a methionine pulse was set resembling the initial methionine concentration of the beginning of the fed-batch.

In parallel, a control fed-batch was run without pulsing methionine. As expected the initial methionine concentration in the fedbatch is quite low with 240 mg L^{-1} (figure 5.14A). The limitation of methionine occurred after 65 h visible by a stop in penicillin production. In the adaptation phase of the pulse higher growth could be observed which switched to penicillin production after around 15 h. Finally, the penicillin concentration could be more than doubled after the pulse (figure 5.14A). On the other hand, the fedbatch without methionine pulse showed no marked increase in penicillin concentration after the limitation (figure 5.14B). Interestingly, no complete depletion of cysteine could be observed for neither of the two cultivations.

The analysis of the generated data with the MFA model illustrates the effect of the pulse on physi-

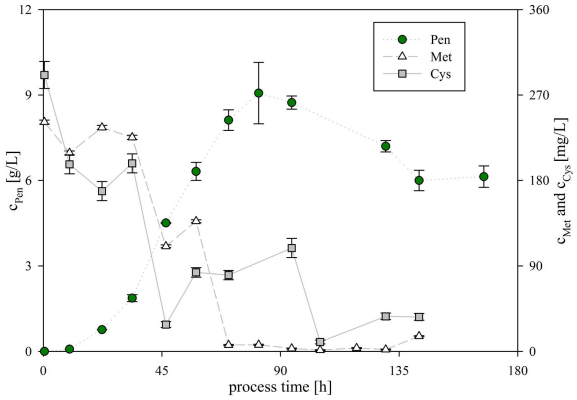

Figure 5.13: Concentrations of penicillin (green circles), methionine (white triangles) and cystine (grey squares) over process time for a standard *Penicillium chrysogenum* fedbatch. The penicillin concentration decreases after 82 h which seems to be linked to the depletion of methionine. Cystine is not completely depleted, but penicillin production still seems to be limited. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

ological parameters. It can be seen that the penicillin/oxygen yield as well as the penicillin/ glucose yield increase significantly after the methionine pulse (figure 5.15). After a renewed decrease of the methionine concentration under 50 mg L^{-1}, a decrease in the penicillin yields can again be observed.

A sensitivity analysis was applied for the generated data in order to analyze a theoretical shift of important substances due to the methionine pulse. Again the results show the highest sensitivity for methionine and cystine (figure 5.16). Hence, although only half of the CSL concentration in the batch media was added, no other raw material ingredient showed critical behavior after extinction of the methionine limitation. Methionine was pulsed but still remains a kMA in this process meaning that sulfur availability still seems to be the most critical path for penicillin formation.

In summary, the hypothesis generated via the MFA model could be experimentally substantiated. The combination of an MFA model and a sensitivity analysis is suitable for the detection of critical media ingredients and the evaluation of complex raw material.

Consideration of enzymatic release from CSL

Bioavailability of amino acids in CSL and enzymatic release kinetics during the process were evaluated previously (unpublished results). Additionally, an effect of consideration of release kinetics on the interpretation of specific amino acid rates could be shown. In total, 10 amino acids showed a marked release in fed-batch. Hence, the calculation of the specific amino acid rates for the MFA model was evaluated with or without consideration of enzymatic release kinetics. For further sensitivity analysis no differences between considering or neglecting release kinetics from CSL were observed. Amino acids with marked release from CSL over the process showed generally low sensitivity on q_{Pen} or m. In this very application changes in bioavailability of substances due to release from CSL seem negligible. Nevertheless, if this method is applied for different raw materials and/ or processes, the raw material peculiarities such as bioavailability should be taken into account.

Discussion

BIOPROCESSES are naturally processes including high variability. In order to guarantee product quality and facilitate high productivity, regulatory guidelines were established by the International Conference on Harmonisation. With respect to raw material, so called critical raw material attributes (CMAs) have to be evaluated in order to allow science-based decisions during raw material quality assessment (Guideline 2009). The proposed method, namely the combination of MFA and sensitivity analysis, fulfills the requirement of a mechanistic understanding of the functional relationships between raw material and process performance. In this study, it has been successfully shown, that this method can be applied for complex raw material, namely CSL, and results in the identification of a kMA, namely methionine. From this point there are different ways of handling this information in order to optimize the process with respect to CSL and productivity:

- The amount of CSL in the batch medium can be

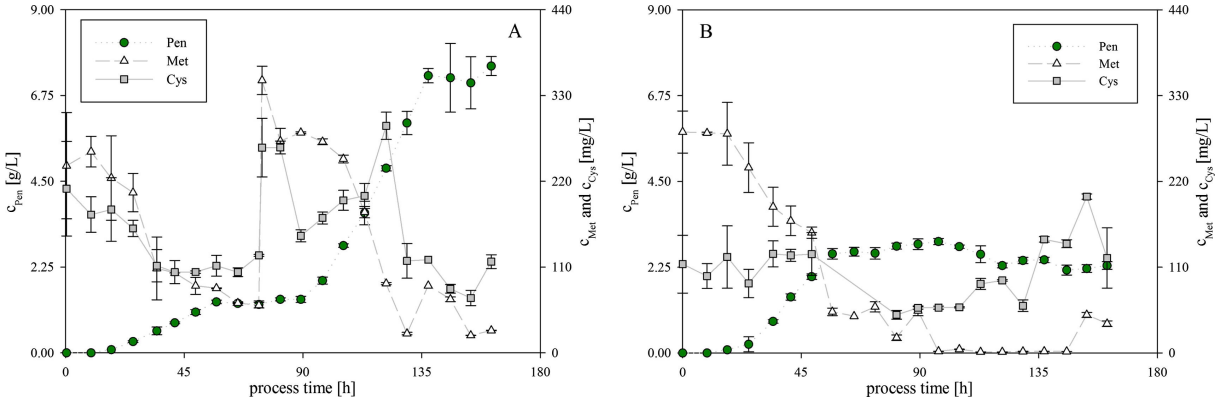

Figure 5.14: Concentrations of methionine, cystine and penicillin for the pulsed fed-batch (A) and the control run (B) over process time. Plot A shows the concentrations before the methionine pulse (0–65 h)) and after the pulse (65–160 h). Continuous uptake of methionine by the fungi can be observed as well as the boost in penicillin production caused by the methionine pulse. Plot B shows the control run. Penicillin production is clearly limited by methionine.

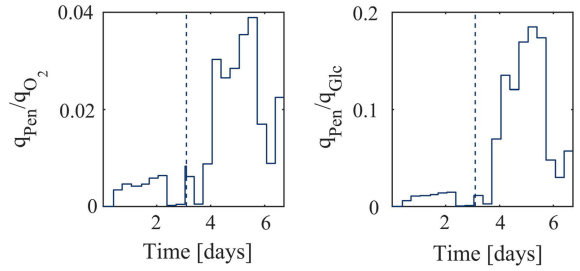

Figure 5.15: Penicillin yields before and after the methionine pulse (dashed line). For both yields an increase can be observed after the pulse, which declines again resulting in an renewed decrease in methionine under 50 mg L^{-1} 1at around 5.3 days.

increased.

- An additional feed (pulse or constant) can be implemented.
- The amount of CSL can be reduced and a feed can be implemented.
- A defined medium can be established.

Generally, complex raw material can cause problems in biopharmaceutical processes. They impair analytics, may add impurities to the process that need consideration in the up and down streaming, and they usually underlie a high lot-to-lot variability leading to process variability. Hence, a reduction of the complex raw material without a boost in costs is favored. As shown in this study, the amount of CSL can be reduced, if an additional pulse of methionine is added, without inducing a loss of productivity but rather producing an increase. An additional feed of methionine could be one solution of a raw material related process improvement. A switch to defined medium could be even more beneficial. However, for this approach all other ingredients of CSL have to be considered as well, especially trace elements and minerals, as they are reported to have an effect on the penicillin process (Posch, Herwig, and Spadiut 2013). The MFA model could be further used for an experimental assessment of a defined medium.

Irrespective of the process and the raw material, we can recommend a workflow for the assessment of CMAs and kMAs in complex raw materials and subsequent process optimization figure 5.17. The composition of complex raw materials is often unknown, and so is the mechanistic understanding between this composition and process performance. Hence, the first step is the establishment of analytical methods that reflect the main ingredients of the raw material, followed by its characterization including raw material peculiarities as bioavailability or miscibility. With the knowledge of the raw material composition, an MFA model can be set up in step 3 includ-

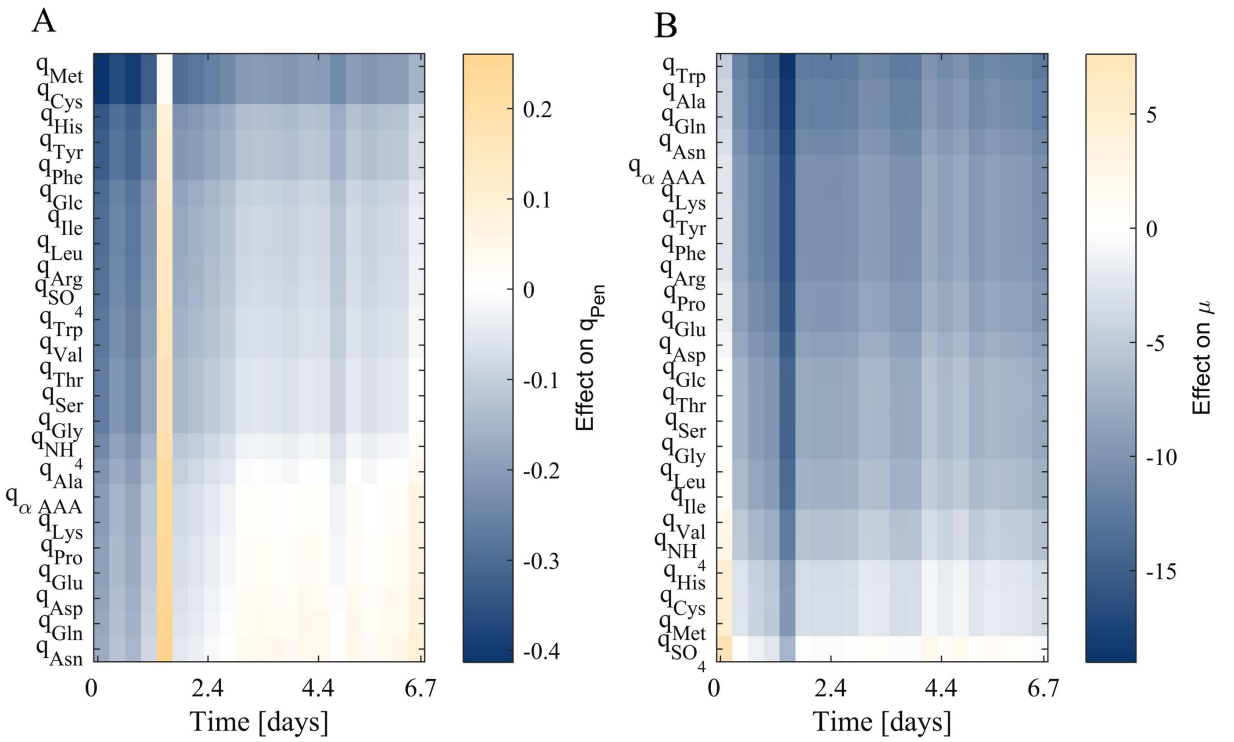

Figure 5.16: Combination of MFA and sensitivity analysis for the generated data including the methionine pulse after 3 days. Again q_{Pen} (plot A) and m (plot B) were chosen as target response; a negative effect indicates a positive effect on the target if the flux is increased. The results did not change markedly compared to the results obtained from the standard cultivation. Hence, irrespective of the methionine pulse and the decrease in CSL, sulfur availability still seems to be the most critical path for penicillin production.

ing raw material relevant fluxes. In order to generate hypotheses on critical raw material attributes a sensitivity analysis is performed in step 4 extracting fluxes that show a correlation with process performance or product quality attributes. These hypotheses have to be experimentally evaluated (step 5). If they cannot be substantiated, the new generated data can be fed in the MFA model and the workflow has to be reentered at step 3. If the experiments substantiate the hypotheses, the process can be adapted according to the gained knowledge. This process adaptation has to be revised by reentering the workflow at step 4 again. If the benefit of the process adaptation is substantiated, an optimized process with respect to raw material attributes can eventually be established.

Additionally to the presented workflow the combination of MFA and sensitivity analysis can be used as platform technology. Irrespective of the process, this combination could be applied for generating process understanding with respect to media composition, raw material quality or metabolic engineering.

Conclusions

QbD for complex raw materials is critical due to their multicomponent matrix. If an in-depth characterization can be achieved, the establishment of functional relations between raw material attributes and process performance or product quality is challenging.

This study introduces a novel method combination of MFA and sensitivity analysis to fill the gap

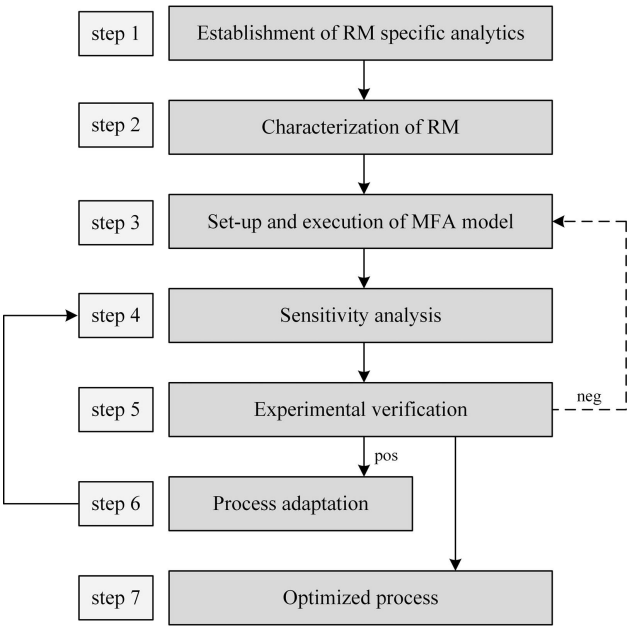

Figure 5.17: Workflow for the assessment of functional relationships between raw material quality and process performance or product quality with the goal of process improvement. After establishment of analytics and characterization of the raw material (RM) an MFA model is set up including RM relevant fluxes. The combination of MFA and sensitivity analysis generates hypotheses about functional relationships, which have to be evaluated experimentally. The process can be adapted and after evaluation of this adaptation by reentering the workflow at step 4, optimized with respect to raw material attributes. If the hypotheses cannot be experimentally substantiated, the workflow has to be iterated at step 3 with the inclusion of the generated data of step 5.

between raw material analysis and process understanding. The proposed method is presented with the complex raw material CSL applied in a *Penicillium chrysogenum* process.

An MFA model was adapted from literature and reduced to the necessary pathways with respect to CSL composition, i.e. mainly amino acid metabolism pathways. Additionally, enzymatic release kinetics of amino acids from CSL during fed-batch cultivation were considered in rate calculations. Sensitivity analysis was applied on the model with respect to q_{Pen} and μ. Methionine could be identified as kMA in CSL for this process.

This hypothesis could be experimentally substantiated by achieving a doubling in penicillin titer after a methionine pulse. It could be shown that MFA combined with sensitivity analysis is a useful tool to generate process understanding with respect to complex raw material. Furthermore, the method allows the identification of raw material variability and the impact of these variances on the process. Finally, we propose a workflow for transferring our findings to other applications. Hence, this combination shows a huge potential for other applications, e.g. in the field of metabolic engineering.

Competing interests

The authors declare that there is no conflict of interest.

Acknowledgements

The authors would like to thank Sandoz GmbH for providing the strain. Additionally, we want to thank Julian Kager for support in the lab and fruitful discussion. Financial support was provided by the Austrian research funding association (FFG) under the scope of the COMET program within the research project “Industrial Methods for Process Analytical Chemistry – From Measurement Technologies to Information Systems (imPACts)” (contract # 843546).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ces.2018.06.075.

References

Agren, R. et al. (2013). "The RAVEN toolbox and its use for generating a genome-scale metabolic model for *Penicillium chrysogenum*". In: *PLoS computational biology* 9.3. DOI: 10.1371%2Fjournal.pcbi.1002980.

Christensen, B. and J. Nielsen (2000). "Metabolic network analysis of *Penicillium chrysogenum* using ^{13}C-labeled glucose". In: *Biotechnology and Bioengineering* 68.6, pp. 652–659. DOI: 10.1002/(SICI)1097-0290(20000620)68:6%3C652::AID-BIT8%3E3.0.CO;2-J.

Douma, R. D. et al. (2010). "Dynamic gene expression regulation model for growth and penicillin production in *Penicillium chrysogenum*". In: *Biotechnology and bioengineering* 106.4, pp. 608–618. DOI: 10.1002/bit.22689.

Goudar, C. T. et al. (2014). "Metabolic flux estimation in mammalian cell cultures". In: *Animal Cell Biotechnology*. Springer, pp. 193–209. DOI: 10.1007/978-1-62703-733-4_13.

Guideline, ICH Harmonised Tripartite (2009). "Pharmaceutical Development Q8 (R2)". In: *Current step* 4.

Henriksen, C. M. et al. (1996). "Growth energetics and metabolic fluxes in continuous cultures of *Penicillium chrysogenum*". In: *Journal of Biotechnology* 45.2, pp. 149–164. DOI: 10.1016/0168-1656(95)00164-6.

Hofer, A. et al. (2018). "Prediction of filamentous process performance attributes by CSL quality assessment using mid-infrared spectroscopy and chemometrics". In: *Journal of biotechnology* 265, pp. 93–100. DOI: 10.1016/j.jbiotec.2017.11.010.

Hönlinger, C. et al. (1988). "Differential effects of general amino acid control of lysine biosynthesis on penicillin formation in strains of *Penicillium chrysogenum*". In: *The Journal of antibiotics* 41.2, pp. 255–257. DOI: 10.7164/antibiotics.41.255.

Jørgensen, H. et al. (1995). "Metabolic flux distributions in *Penicillium chrysogenum* during fed-batch cultivations". In: *Biotechnology and Bioengineering* 46.2, pp. 117–131. DOI: 10.1002/bit.260460205.

Kleijn, R. J. et al. (2007). "Cytosolic NADPH metabolism in penicillin-G producing and non-producing chemostat cultures of *Penicillium chrysogenum*". In: *Metabolic engineering* 9.1, pp. 112–123. DOI: 10.1016/j.ymben.2006.08.004.

Kroll, P. et al. (2017). "Workflow to set up substantial target-oriented mechanistic process models in bioprocess engineering". In: *Process Biochemistry* 62, pp. 24–36. DOI: 10.1016/j.procbio.2017.07.017.

Lanan, Maureen (2009). "QbD for raw materials". In: *Quality by Design for Biopharmaceuticals: principles and case studies*, pp. 193–209. DOI: 10.1002/9780470466315.ch11.

Liggett, R. W. and H. Koffler (1948). "Corn steep liquor in microbiology". In: *Bacteriological reviews* 12.4, p. 297.

Nielsen, J. (1997). Physiological engineering aspects of *Penicillium chrysogenum*. World Scientific. DOI: 10.1142/3195.
While being short and concise, it is even today one of the most useful books for everyone working with *Penicillium chrysogenum*.

Posch, A. E., C. Herwig, and O. Spadiut (2013). "Science-based bioprocess design for filamentous fungi". In: *Trends in biotechnology* 31.1, pp. 37–44. DOI: 10.1016/j.tibtech.2012.10.008.

Posch, A. E., C. Koch, et al. (2013). "Combining light microscopy, dielectric spectroscopy, MALDI intact cell mass spectrometry, FTIR spectromicroscopy and multivariate data mining for morphological and physiological bioprocess characterization of filamentous organisms". In: *Fungal genetics and biology* 51, pp. 1–11. DOI: 10.1016/j.fgb.2012.11.008.

Pyle, AJH (1954). "Some aspects of the amino acid metabolism of *Penicillium chrysogenum*". In: *Microbiology* 11.2, pp. 191–194. DOI: 10.1099/00221287-11-2-191.

Sin, G., K. V. Gernaey, and A. E. Lantz (2009). "Good modeling practice for PAT applications: propagation of input uncertainty and sensitivity analysis". In: *Biotechnology progress* 25.4, pp. 1043–1053. DOI: 10.1002/btpr.166.

Van Gulik, W. M. et al. (2000). "Application of metabolic flux analysis for the identification of metabolic bottlenecks in the biosynthesis of penicillin-G". In: *Biotechnology and Bioengineering* 68.6, pp. 602–618. DOI: 10.1002/(SICI)1097-

0290(20000620)68:6%3C602::AID-BIT3%3E3.0.
CO;2-2.

Part III

Conclusion and Outlook

Chapter 6

Conclusion

> Die Zusammenhänge sind deutlich, wenn ich auch nicht weiß, welche Zusammenhänge.
>
> Günter Eich

Summary

METABOLIC flux analysis could be considered an undervalued and underutilised method in biotechnology. In contrast to other mathematical methods its formalism is easily accessible and interpretable to the biologist, and in principle the metabolome is well understood and textbooks and databases for a wide array of organisms are widely available. Even without using real data, it can be helpful to investigate a system of interest and get a clearer picture of its implications and how changes in inputs or model formulation impact its predictions.

Within this thesis the following new knowledge could be established:

- Estimates for release kinetics of bound amino acids in corn steep liquor during penicillin fermentation were determined. It was demonstrated that release kinetics are different for different amino acids and that the neglect of release kinetics can lead to misleading estimates of uptake rates.
- It could be shown that the effect of corn steep liquor extends into the fed-batch phase as it is not completely depleted in the batch phase. One amino acid, valine, was even only released from its bound pool in the fed-batch phase.
- By combining metabolic flux analysis with sensitivity analysis we could in principle detect changing limitations and demands during the process. Limitations could be interpreted on a mechanistic basis by investigating the change in estimated internal fluxes.
- It could be shown that NADPH production is limiting the yield of penicillin. By adding sulphur containing amino acids the NADPH demanding sulphur utilization can be circumvented and NADPH can be used for other essential steps in penicillin production. These results are in accordance to established knowledge (Van Gulik et al. 2001).

Caveats

Nonetheless, the use of metabolic flux analysis and its role described here, should be seen as in aiding the mostly qualitative interpretation of experimental results and helping to generate new testable hypothesis. It is in this form not suitable for other tasks, such as process control or simulation. Such a use would not only demand a much more elaborate set of experiments and the determination of regulatory and kinetic information, but also the verification and validation of the stoichiometric model through ^{13}C-labelling experiments. Especially the last part is to-

date still a demanding task that is not as accessible as classical metabolic flux balancing and can lead to misinterpretation of data all too easily (Winden, Verheijen, and Heijnen 2001).

References

Van Gulik, W. M. et al. (2001). "Energetics of growth and penicillin production in a high-producing strain of *Penicillium chrysogenum*". In: *Biotechnology and bioengineering* 72.2, pp. 185–193. DOI: 10.1002/1097-0290(20000120)72:2%3C185::AID-BIT7%3E3.0.CO;2-M.

Winden, W. van, P. Verheijen, and S. Heijnen (2001). "Possible pitfalls of flux calculations based on ^{13}C-labeling". In: *Metabolic engineering* 3.2, pp. 151–162. DOI: 10.1006/mben.2000.0174.

Chapter 7

Outlook

> Nil actum credens cum quid superesset agendum.
>
> Nothing is done, if something is left to do.
>
> Marcus Annaeus Lucanus

DESPITE the many advances in biotechnology over the last decades there is no shortage of pressing scientific problems for the academic and industrial community. The following paragraphs are meant to give a short overview of related topics that might be worthy of further research interest and might lead to the establishment of new valuable approaches to bioprocess engineering.

Towards dynamic metabolic flux analysis

In its essence, metabolic flux analysis assumes a system at steady state and disregards any regulatory mechanisms which is at odds with both conditions encountered in the environment as in an industrial setting. Therefore, it can not be used for simulation which limits its application to engineering. Methods to circumvent this limitation are the combination of MFA with kinetic models (Naderi et al. 2011), or the combination of pathway analysis with kinetic models (Provost and Bastin 2004). While most often the extracellular rates are described by kinetic rates in many cases intracellular reactions are the driving force behind internal metabolism and external uptake. Therefore Nolan and Lee (2011) used kinetic expressions for the intracellular rates they saw as the central fluxes of the system. However, still many challenges remain to be overcome to achieve dynamic metabolic flux analysis, especially in regards to ^{13}C labelling experiments where incorporation of labelled substrates is relatively slow (Ahn and Antoniewicz 2012). To the author's knowledge the use of different stoichiometric matrices for different process phases as a method to model shifts in metabolism has so far not been investigated in the literature but might be an idea worth pursuing.

Another way of achieving dynamic metabolic flux analysis would be by potential advances in *Process Analytical Technologies*. If sufficient on-line measurements of rates can be made available, either by direct measurements or through soft-sensors, one can determine intracellular fluxes through MFA on-line in real time (Goudar et al. 2006).

Beyond the (right) null space

MFA concerns itself with exploring the pseudo-steady state, i.e., the null space that is described by $G^T v = 0$. However, under changing conditions intracellular metabolite concentrations can be subject to change and play a vital role in regulatory mechanisms. As an illustrative example Fungi can also use their vacuoles for storage of metabolites (Veses, Richards, and Gow 2008). In this case the pseudo-steady state is not really violated, as the intracellular concentrations of soluble metabolites do not

change, but models need to be able to appropriately describe these vacuole pools. Other ways of thinking beyond this very small subspace of possible solutions has led to the definition of metabolite pools as the left null space of G^T (Famili and Palsson 2003). Other constraints imposed on the stoichiometric matrix might similarly yield new important concepts. An especially interesting approach that yields dynamic cell models that make prediction also outside the pseudo-steady state has been termed *mass action stoichiometric simulation* (Mostolizadeh, Dräger, and Jamshidi 2019). It combines data from genomics and metabolomics with thermodynamic and other constraints to calculate kinetic parameters at steady state, so that behaviour can be simulated for any conditions.

Segregation in metabolic flux analysis

Tsuchiya, Fredrickson, and Aris (1966) have been credited with the classification of models into structured and non-structured and segregated and non-segregated models. While this classification is very simplistic and leaves out many other important aspects, it has proven to be very useful in clearly communicating basic assumptions (see the figure by Bailey (1998)). However, while modelers have started using structured models in metabolic flux analysis from an early point on, the use of segregated models is still quite rare. While the need for segregated flux models has been acknowledged (Dauner 2010), the author could only find one publication that uses truly segregated flux analysis (Pardelha et al. 2013). Even publications that already utilize segregated models for growth fall back on non-segregated models for metabolic flux analysis (Wahl et al. 2008). The exploration of this idea will not only demand new mathematical methodologies but also new analytical techniques for experimental verification.

Towards mechanistic media optimization

While a recent review by Singh et al. (2017) describes several modern statistical approaches to media optimization including genetic algorithms, artificial neural networks, and Nelder-Mead simplex, no mechanistic approaches were mentioned. This lack of mechanistic methods in media optimization is concerning as optimal media composition can not only vary between species and strains but even between clones of the same cell line (Pan et al. 2017). As a consequence, better and faster methods than currently available are needed to ensure optimal process conditions.

Galbraith et al. (2018) propose a two step method for future media optimization: At first, the optimal media components are defined by the use of a stoichiometric model, while in the second step the optimal media concentrations are determined through kinetic simulations. Progress has been made in this direction by Robitaille, Chen, and Jolicoeur (2015) who could show that one parametrization of such a kinetic-metabolic model was able to predict metabolic behaviour on two different media. Still, much work is left to be down as the method described by Robitaille, Chen, and Jolicoeur (2015) does not scale well and is therefore not able to explore the potentials of completely new media as proposed by Galbraith et al. (2018).

Additionally, during media optimization often only upstream processing is considered, ignoring the effects on product purification and polishing. In the future, a more holistic approach might become the new paradigm, where processes are designed and optimized *in silico* not seperately per each unit operations but as a whole.

The future of penicillin production

Despite β-lactam antibiotics constituting 40 % of the antibiotic market (Kresse, Belsey, and Rovini 2007), the high competitiveness of it has resulted in reduced investment into new innovations in peni-

cillin production (Luepke et al. 2017). However, as the problem of antibiotic resistances becomes more pressing, necessity and political pressure might lead to an increased research interest in the near future and the development of radically new approaches to penicillin production.

Even though only filamentous fungi produce penicillins naturally, through the means of synthetic biology the pathway for its biosynthesis could be transferred to the yeasts *Ogataea (Hansenula) polymorpha* (Gidijala et al. 2009) and *Saccharomyces cerevisiae* (Awan et al. 2017). While the penicillin concentrations achieved are two to three orders of magnitude below industrial titers, they might still be an attractive alternative production host. Yeasts not only have smaller and better annotated genomes (Cairns et al. 2019), some transformation techniques such as homologous recombination are also more efficient in yeasts (Wang 2007).

Another alternative to established production would be to eliminate the need of biological producers altogether. The first total synthesis of a penicillin was achieved in the late 50s by Sheehan and Henery-Logan (1959), however early hopes of chemical synthesis on an industrial scale were disappointed. Bycroft and Shute (1987) state that "it is likely that fermentation will remain the predominant source of penicillins for the projectable future" and have proven right so far. However, often it is only the need for unstable and expensive cofactors such as ATP or NAD(P) that prevents the application of in-vitro biosynthesis. Recent alternative regeneration systems for oxidoreductases have emerged with a clear vision of synthetic application (Zhang and Hollmann 2018). Further advances in this field might eventually lead to membrane bound enzymes to replace microorganisms for many established bioprocesses such as penicillin production.

References

Publications of special significance and value have their titles highlighted in boldface and have been complemented by an explanatory comment.

Ahn, W. S. and M. R. Antoniewicz (2012). "Towards dynamic metabolic flux analysis in CHO cell cultures". In: *Biotechnology journal* 7.1, pp. 61–74. DOI: 10.1002/biot.201100052.

Awan, A. R. et al. (2017). "Biosynthesis of the antibiotic nonribosomal peptide penicillin in baker's yeast". In: *Nature communications* 8.1, pp. 1–8. DOI: 10.1038/ncomms15202.

Bailey, J. E. (1998). "**Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities**". In: *Biotechnology progress* 14.1, pp. 8–20. DOI: 10.1021/bp9701269.
This paper gives a good overview over the history of modelling in bioengineering and even after more than 20 years provides useful input for the prospective modeller. Additionally, this publications deals with a few all to common misconceptions such as "shear damage" and what really makes growth factors more effective.

Bycroft, B. W. and R. E. Shute (1987). "Chemistry and biosynthesis of penicillins and cephalosporins". In: *Penicillium and Acremonium*. Springer, pp. 113–160. DOI: 10.1007/978-1-4899-1986-1_5.

Cairns, T. C. et al. (2019). "Moulding the mould: understanding and reprogramming filamentous fungal growth and morphogenesis for next generation cell factories". In: *Biotechnology for biofuels* 12.1, p. 77. DOI: 10.1186/s13068-019-1400-4.

Dauner, M. (2010). "From fluxes and isotope labeling patterns towards in silico cells". In: *Current opinion in biotechnology* 21.1, pp. 55–62. DOI: 10.1016/j.copbio.2010.01.014.

Famili, I. and B. Ø. Palsson (2003). "**The convex basis of the left null space of the stoichiometric matrix leads to the definition of metabolically meaningful pools**". In: *Biophysical journal* 85.1, pp. 16–26. DOI: 10.1016/S0006-3495(03)74450-6.
The principles of extreme pathway analysis are not applied to determine pathways from reactions, but conserved metabolite pools from the metabolites.

Galbraith, S. C. et al. (2018). "Media formulation optimization: current and future opportunities". In: *Current opinion in chemical engineering* 22, pp. 42–47. DOI: 10.1016/j.coche.2018.08.004.

Gidijala, Loknath et al. (2009). "An engineered yeast efficiently secreting penicillin". In: *PLoS One* 4.12. DOI: 10.1371/journal.pone.0008317.

Goudar, C. et al. (2006). "Towards industrial application of quasi real-time metabolic flux analysis for mammalian cell culture". In: *Cell Culture Engineering*. Springer, pp. 99–118. DOI: 10.1007/10_020.

Kresse, H., M. J. Belsey, and H. Rovini (2007). *The antibacterial drugs market*. DOI: 10.1038/nrd2226.

Luepke, K. H. et al. (2017). "Past, present, and future of antibacterial economics: increasing bacterial resistance, limited antibiotic pipeline, and societal implications". In: *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy* 37.1, pp. 71–84.

Mostolizadeh, R., A. Dräger, and N. Jamshidi (2019). "Insights into Dynamic Network States Using Metabolomic Data". In: *High-Throughput Metabolomics*. Springer, pp. 243–258. DOI: 10.1007/978-1-4939-9236-2_15.

Naderi, S. et al. (2011). "Development of a mathematical model for evaluating the dynamics of normal and apoptotic Chinese hamster ovary cells". In: *Biotechnology progress* 27.5, pp. 1197–1205. DOI: 10.1002/btpr.647.

Nolan, R. P. and K. Lee (2011). "Dynamic model of CHO cell metabolism". In: *Metabolic engineering* 13.1, pp. 108–124. DOI: 10.1016/j.ymben.2010.09.003.

Pan, X. et al. (2017). "Selection of chemically defined media for CHO cell fed-batch culture processes". In: *Cytotechnology* 69.1, pp. 39–56. DOI: 10.1007/s10616-016-0036-5.

Pardelha, F. et al. (2013). "Segregated flux balance analysis constrained by population structure/function data: the case of PHA production by mixed microbial cultures". In: *Biotechnology and bioengineering* 110.8, pp. 2267–2276. DOI: 10.1002/bit.24894.

Provost, A. and G. Bastin (2004). "**Dynamic metabolic modelling under the balanced growth condition**". In: *Journal of Process Control* 14.7, pp. 717–728. DOI: 10.1016/j.jprocont.2003.12.004.
A detailed description of the general dynamic metabolic model, metabolic flux analysis, and how to generate macro-reactions from the stoichiometric model. Uses simple Monod kinetics for dynamic modelling of the macro-reactions, which were dereived from pathway analysis by using elementary flux modes. Based on the assumptions of balanced and exponential growth.

Robitaille, J., J. Chen, and M. Jolicoeur (2015). "A single dynamic metabolic model can describe mAb producing CHO cell batch and fed-batch cultures on different culture media". In: *PloS one* 10.9. DOI: 10.1371/journal.pone.0136815.

Sheehan, J. C. and K. R. Henery-Logan (1959). "The total synthesis of penicillin V". In: *Journal of the American Chemical Society* 81.12, pp. 3089–3094.

Singh, V. et al. (2017). "Strategies for fermentation medium optimization: an in-depth review". In: *Frontiers in microbiology* 7, p. 2087. DOI: 10.3389/fmicb.2016.02087.

Tsuchiya, H. M., A. G. Fredrickson, and R. Aris (1966). "Dynamics of microbial cell populations". In: *Advances in chemical engineering*. Vol. 6. Elsevier, pp. 125–206. DOI: 10.1016/S0065-2377(08)60275-6.

Veses, V., A. Richards, and N. A. R. Gow (2008). "Vacuoles and fungal biology". In: *Current opinion in microbiology* 11.6, pp. 503–510. DOI: 10.1016/j.mib.2008.09.017.

Wahl, A. et al. (2008). "Metabolic flux model for an anchorage-dependent MDCK cell line: Characteristic growth phases and minimum substrate consumption flux distribution". In: *Biotechnology and Bioengineering* 101.1, pp. 135–152. DOI: 10.1002/bit.21873.

Wang, X. (2007). "Effect of small rank modification on the condition number of a matrix". In: *Computers & Mathematics with Applications* 54.6, pp. 819–825. DOI: 10.1016/j.camwa.2006.05.031.

Zhang, W. and F. Hollmann (2018). "Nonconventional regeneration of redox enzymes–a practical approach for organic synthesis?" In: *Chemical communications* 54.53, pp. 7281–7289. DOI: 10.1039/C8CC02219D.